Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Bayesian prediction of glacial discharge in Antarctica using copulas

Autor:

Mario Gómez Díaz

Director/es:

Mª Concepción Ausín

Mª del Carmen Domínguez

DEPARTAMENTO DE ESTADÍSTICA

Getafe, 22 de diciembre de 2017

Universidad
Carlos III de Madrid
www.uc3m.es

(a entregar en la Oficina de Posgrado, una vez nombrado el Tribunal evaluador , para preparar el documento para la defensa de la tesis)

TESIS DOCTORAL

Bayesian prediction of glacial discharge in Antarctica using copulas

Autor: *Mario Gómez Díaz*

Director/es: Mª Concepción Ausín

Mª del Carmen Domínguez

Firma del Tribunal Calificador:

Firma

Presidente: Juan Miguel Marín Diazaraque

Vocal: Patrícia Cortés de Zea Bermudez

Secretario: Ana Justel Eusebio

Calificación:

Getafe, 22 de diciembre de 2017

Universidad Carlos III

PH.D. THESIS

Bayesian prediction of glacial discharge in Antarctica using copulas

Author:

Mario Gómez Díaz

Advisor:

M. Concepción Ausín & M. Carmen Domínguez

DEPARTMENT OF STATISTICS

Getafe, Madrid, December 22, 2017

To my wife and my daughters!

Acknowledgements

A mi mujer, Gely, que siempre cree en mi más que yo mismo, que sabe pincharme cuando no me salen las cosas para que no tire la toalla. Ella fue la primera en empujarme hacia un proyecto que me asustaba por la magnitud que suponía. Por aceptar con buena cara las ausencias, los viajes, los agobios, . . . , y sobre todo por este último tramo que ha resultado tan duro.

A mis hijas, Pilar y Laura, aunque no lo sepan, ellas me han dado fuerzas para seguir adelante, poder servirles de ejemplo, de que con esfuerzo se puede conseguir llegar a cualquier meta, ha sido un buen revulsivo en los momentos difíciles. Ellas son capaces de arrancarme una sonrisa en cualquier momento con sus locuras, de sacarme de la rutina para contarme sus cosas.

A mis tutoras, a Conchi por confiar en mi para la realización de esta tesis, por darme todo su apoyo y confianza. Por guiarme en todo este proceso. Sus comentarios, tanto académicos como de futuro han sido muy provechosos. A Karmenka, por hacerme partícipe de su fantástico proyecto, por transmitirme el entusiasmo y confiar en mi para aprovechar el gran trabajo de campo que ella y Adolfo están realizando desde GLACKMA.

A mis padres, que desde bien pequeño me transmitieron las ganas de aprender, por proporcionarme las herramientas para entender el mundo y enseñarme la importancia que tiene estudiar y aprender cosas nuevas cada día.

A mis hermanos que, aun siendo más pequeños que yo, me precedieron en el camino de perseguir los sueños. Ellos consiguieron sus objetivos con gran esfuerzo y me demostraron que yo también podía conseguirlo.

A Angelita, que no ha podido verme terminar este proyecto, pero estoy seguro que está muy

orgullosa de saber que lo he conseguido, en sus momentos más duros seguía teniendo palabras de aliento para mí. A Nisio y familia que me acogieron en su casa cada vez que tenía que venir a Madrid, siempre pendientes que no me faltara nada.

A Loren que siempre ha creído en mi potencial y que me puso en contacto con la Universidad Carlos III para que me guiaran en este proyecto. Las charlas que hemos tenido estos últimos años han sido muy provechosas para seguir adelante.

A mis profesores de matemáticas, en especial a Rufino y Gene que despertaron mi interés por las matemáticas desde bien pequeño, a Carlos y Pilar que en el instituto alimentaron mi necesidad de conocer más de esta Ciencia y a los profesores de la Facultad de Matemáticas de Salamanca que me enseñaron a ver el mundo desde la óptica de las matemticas.

A mis amigos de Puente, ellos saben estar siempre presentes, aunque pasemos tiempo sin vernos. Esas reuniones que te hacen darte cuenta que los problemas no son tan grandes como parecen y que vivir despacio es la mejor manera de disfrutar de la vida.

A mis compañeros de departamento, en especial a María, que siempre ha estado pendiente de que mi estancia en la Universidad haya sido lo más agradable posible, ayudándome a integrarme y acompañándome en mis visitas a la Universidad. A Hoang y a Javi por acompañarme en el despacho, aunque no hemos compartido demasiado tiempo juntos, siempre han estado dispuestos a echar una mano. A Joao, Antonio, Ángela, Nico, Francisco, Diego, Ginette, Elisa. A los profesores del Departamento de Estadística de la Universidad Carlos III que me han ayudado tanto en mi formación como en la docencia.

Abstract

Glaciers are considered sensors of the Global Warming. The study of their mass balance is essential to understand their future behaviour. One of the components of this mass balance is the loss of water produced by melting, this is known as the glacier discharge. The aim of this work is to analyse the relationship among the glacier discharge and other meteorological variables such as temperature, humidity, solar radiation and precipitation, and to find a model that allow us to forecast future values of the glacier discharge. **In Chapter 2**, we propose the use of time-varying copula models for analysing the relationship between air temperature and glacier discharge, which is clearly non constant and non-linear through time. A bivariate copula model is defined, where both, the marginal and copula parameters, vary periodically along time; following a seasonal dynamic. Full Bayesian inference is performed such that the marginal and copula parameters are estimated in a one single step, in contrast with the usual two-step approach. Bayesian prediction and model selection are also carried out for the proposed model such that Bayesian credible intervals can be obtained for the conditional glacier discharge given a value of the temperature at any time point. **In Chapter 3**, as a second model, a vine copula structure is proposed to model the multivariate and nonlinear dependence among the glacier discharge and the other related meteorological variables. The multivariate distribution of these variables is divided in four cases according to the presence or not of positive discharge and/or positive precipitation. Then, each different case is modelled with a vine copula. Seasonal effects in this second model are captured by using different parameters for each season. The conditional probability of zero discharge for given meteorological conditions is obtained from the proposed joint distribution. Moreover, the structure of the vine copula allows us to derive the

conditional distribution of the glacier discharge for the given meteorological conditions. Three different prediction methods are used and compared for the future values of the discharge. In order to improve the second model, **Chapter 4** proposes a hierarchical structure where the relationships between the meteorological variables in each season and in each case are led by common hyperparameters. Bayesian inference is performed over the hierarchical structure with the help of Approximate Bayesian Computation (ABC) techniques.

All the proposed methodologies are applied to a large data base collected since 2002 by the GLACKMA association from a measurement station located in the King George Island in the Antarctica which records values of the liquid discharge from the Collins glacier.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Glaciers

The study of the mass balance in glaciers is crucial for the accurate quantification of water resources (Hamlet and Lettenmaier (1999); Marsh (1999)). Mass balance is defined as the difference between accumulation (mainly the fallen snow) and ablation which includes processes such as sublimation, calving and melting. In particular, most of the liquid water is lost by surface melting and runoff and surface melting, percolating inside the glacier and exit by the front or the base. This is known as glacier discharge and it can be defined as the rate of flow of meltwater through a vertical section perpendicular to the direction of the flow (Cogley et al. (2011)).

The ice melting on the ice cap of the glacier drops into the glacier through the moulins, that are vertical or nearly vertical shafts (see Fig. 1.1). This liquid water flows through the glacier and melts the inner ice by friction, increasing the amount of melted water. This flow creates a network of pipes, galleries and caverns called cryokarst, as the one we can see in Fig. 1.2. When this water breaks the glacier slope, flows out and produces the glacier discharge.

Glaciers can be classified in three different main types in terms of their thermal condition (Baranowski and Jurasz (1977); Eraso and Pulina (1994)). First, the polar glaciers where the ice is always below freezing point from the surface to its bedrock. Second, the subpolar glaciers

Figure 1.1: Moulin: A deep shaft, nearly vertical and of roughly circular cross section, formed when surface meltwater enlarges a crack in the ice by transferring kinetic and thermal energy to its walls (Cogley et al. (2011)).

Figure 1.2: Pipe of a cryokarst with the liquid water flowing inside the glacier. This water create the criokarst by melting friction.

include both ice at freezing point and at below temperature. And third, temperate glaciers which are around freezing point throughout the year. The polar glaciers have no cryokarst whereas the subpolar and the temperate glaciers have.

Modelling glacier discharge is a quite important issue in climate and hydrology research (Jansson et al. (2003); La Frenierre and Mark (2014)). An extensive review of the different approaches for glacier melt modelling can be found in Hock (2005). These models are usually classified in two main categories: energy balance models, which evaluate the most important energy fluxes between the atmosphere and the glacier surface. These fluxes are computed from physically based calculations (Braun (2001); Sicart et al. (2008)) and try to solve these complex equations and measurements relating the gain and loss of ice in glacier systems (Ohmura (2001); Willis et al. (2002)). Alternatively, temperature index models use only the air temperature to empirically model this relationship. A complete review of temperature index methods can be found in Hock (2003). A leading form of temperature-index models is the so called degree-day model which is based on an assumed linear relationship between ablation and air temperature, usually expressed in the form of positive temperature sums. There are some studies that incorporate more variables to this model, such as the direct solar radiation (Hock (1999)) or the albedo and the shortwave radiation (Pellicciotti et al. (2005)). The main problem in this type of models is that they implicitly assume a linear relationship between temperature and discharge, which is not realistic in practice. Also, these models only take into account the ice melting on the surface but not the one produced inside by friction.

GLACKMA, whose name is an acronym of *GLAciares, CrioKarst y Medio Ambiente* (Glaciers, Cryokarst and Environment), is an association which promotes scientific research in the polar regions, see http://www.glackma.org. Their researchers have been developing a project since 2001 with the aim of using the glaciers as natural warming sensors, as explained e.g. in Hock (2005) and Bers et al. (2013). GLACKMA has implemented eight stations, called Pilot Experimental Catchment areas (*CPE - Cuenca Piloto Experimental*) which are working continuously to register glacier discharge hourly values. Fig. 1.3 shows the location of the eight CPE. This network includes different latitudes in both hemispheres, which allow a comparative control of the glacier discharge according to the evolution of the global warming. There are four CPE in the North Hemisphere: CPE-TAR-68 N (Swedish Arctic), CPE-KVIA-64 N (Iceland), CPE-ALB-79 N (Svalbard) and CPE-OBRU-68 N (Northern Ural mountains) and four CPE in the South Hemisphere: CPE-KG-62 S (Insular Antartic), CPE-HUE-49 S (Patagonia Argentine), CPE-ZS-51 S (Patagonia Chilean) and CPE-VER-65 S (Antarctic Peninsula). The CPE located at Svalbard, Insular Antarctica and Antarctic Peninsula are installed in subpolar glaciers whereas the other CPE are placed in temperated glaciers. There are two of these CPE that share the same latitude, CPE-TAR-68 N and CPE-OBRU-68 N, but they are at different altitude, 1200m and

450m respectively, in order to study the effect of the altitude in the glacier discharge. GLACKMA obtains 8670 records per year for the glacier discharge and other meteorological variables for each one of these stations. The first station, CPE-ALB-79 N, was installed in 2001.

In this thesis, we will concentrate on one of these stations, namely CPE-KG-62 S, located at the King George Island near to the Antarctic Peninsula, where there is available glacier discharge data since 2002. The Antarctic Peninsula is considered as one of the Recent Rapid Regional Climate Warming, which refer to those areas where the regional changes, due to the effect of the Global Warming, have been deeper than the worldwide mean, as noted by the Intergovernmental Panel on Climate Change (IPCC) (Turner et al. (2005); Vaughan et al. (2002)). Periods of melting of the glaciers in this area have been increasing year by year (Domínguez and Eraso (2007)). As consequence, there has been a retreat of the glaciers and changes in their heights (Rückamp et al. (2011)). Also, trends in surface melting have been found (Barrand et al. (2013)). King George is the largest of the South Shetland Islands, which is an archipelago placed near of the coast of the Antarctic Peninsula in the Southern Ocean. Collins Glacier, with around km , covers most of the King George Island except the south-western end of the island, where the Fildes Peninsula is located. See Braun (2001) for a complete description of the island.

More specifically, the GLACKMA measuring station is installed in a canyon near the Uruguayan Base Artigas (62 11’03”S, 58 54’41”W). The study area is located at the SSW side of icecap Collins, known as Smaller Dome or Bellinsghausen Dome. Collins Glacier has almost all of King George island surface. The hydrological discharge creates a fluvial network set up by nine springs that flow into both sides of the coast. The catchment area is located in the south slope that has five springs that discharge their water into a small lake. This lagoon flows into the sea by a single stream, the measuring station CPE-KG-62 S is located in this stream. These five springs drain water from a catchment area with a total surface of km , which comprises km of glacier surface, km of peripheral moraine and km of fluvial surface.

The glacier discharge is calculated from the river level using a calibration function which was determined in several campaigns of gaugings (Domínguez (2004)). In order to compare the

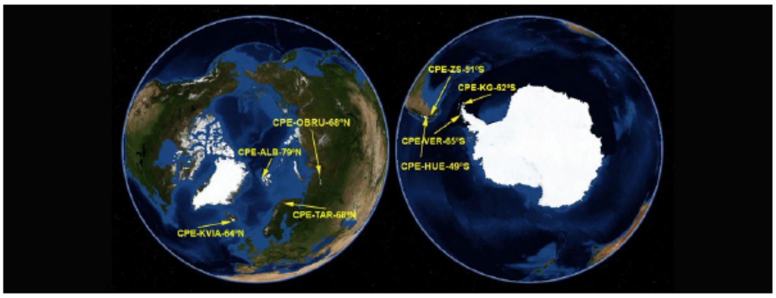

Figure 1.3: Location of the eight Experimental Catchment Areas that GLACKMA has implemented.

measurements of different CPE, the glacier discharge is divided by the surface of the glacier icecap to obtain the specific glacier discharge that is measured in . For a detailed description see Domínguez et al. (2004) and Domínguez and Eraso (2007).

1.2 Study area and data description

The GLACKMA monitoring station was installed in January of 2002 and consisted of a sounder with sensors for water temperature, conductivity and river level. After two years of hourly registrations, the hard meteorological conditions during the austral winter in 2003 caused a series of invalid records. A new high-quality sounder was then set up which, although it only registers values from the river level, it is much more resistant under extreme conditions. The glacier discharge, measured in , can be accurately estimated as an exponential function of the river level using a classical regression fit with (Domínguez and Eraso (2007); Gómez et al. (2017)). Also, we have selected a collection of meteorological variables as covariates.

T: Mean daily air temperature .

H: Mean daily percentage of humidity ( ).

R: Mean daily solar radiation ( ).

P: Daily accumulated precipitation ( )

D: Mean daily specific glacier discharge ( ).

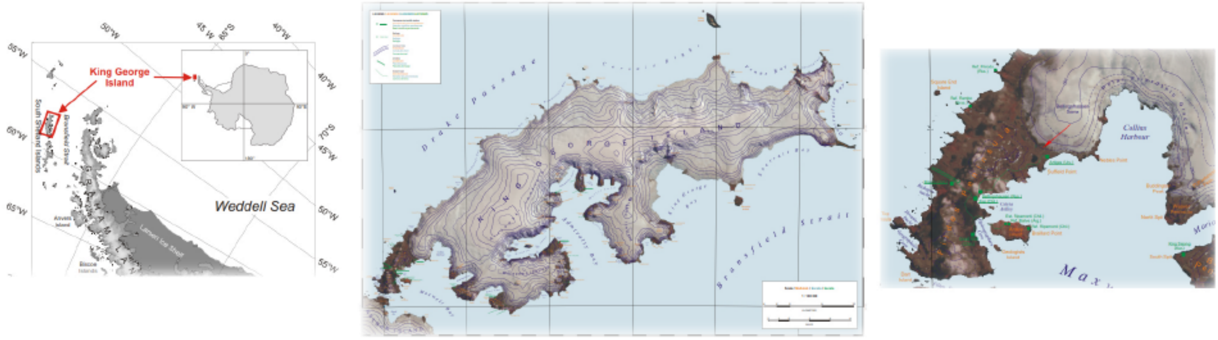

Figure 1.4: Left panel shows the location of King George Island in the Antarctica. Central panel shows the island, mostly covered by Collins glacier. Right panel zooms in the location of the CPE-KG-62 S station, indicated with a red arrow. See Braun et al. (2002).

These meteorological data have been provided, via GLACKMA, by the Bellinghausen Russian base located at 4 km from the CPE-KG-62 S station. Fig. 1.4 shows the study area, located in the south west side of the King George island, where GLACKMA has placed this Pilot Experimental Catchment Area. The available data for this project are from October 1st 2002 to September 30th 2012, with a missing data year for the records of the discharge, from December 1st 2003 to June 30th 2004 (213 consecutive missing values).

A preliminary study of these data shows that discharge and precipitation have a large number of zero-values. In particular, the value of the discharge was zero in of the observed days, because of the fact that the CPE is in a subpolar glacier that only has discharge from late spring to late autumn. The value of the precipitation was zero in of the observed days. This fact will be a definite impact in the design of the different models. Table 1.1 shows the summary statistics and Fig. 1.9 the histograms of all the variables. Observe that the histograms are left-skewed, except from the pressure and humidity. The solar radiation has a large standard deviation mainly because of the long duration of the nights in the Austral Winter (from June 21st to September 21st) contrasted with the long duration of the days in the Austral Summer (from December 21st to March 21st). Discharge and precipitation are non-negative random variables, radiation is a positive random variable and humidity is defined from zero to 100.

Now, we obtain the means of each variable, for each hydrological year, which are defined

	TEMP	HUMI	RADI	PREC	DISC
Min.	-23.400	51.750	0.208	0.000	0.000
1stQu.	-4.061	83.710	5.291	0.000	0.000
Median	-1.120	89.540	22.542	0.700	0.000
Mean	-2.216	88.520	31.488	1.889	0.057
3rdQu.	0.620	94.500	50.0000	2.400	0.056
Max.	5.783	100.000	142.083	43.500	1.306
sd	4.079	7.378	30.155	3.097	0.121
Skewness	-1.286	-0.690	1.0436	3.699	3.527
Kurtosis	1.784	0.0540	0.423	22.198	18.154

Table 1.1: Summary statistics for the meteorological variables.

YEAR	TEMP	HUMI	RADI	PREC	DISC	Days with discharge	Days without precipitation
2002/2003	-2.1146	89.0945	32.9772	1.2479	0.0415	91	127
2003/2004	-1.8650	88.6885	33.9471	1.2541	NA	NA	128
2004/2005	-2.0430	89.3017	31.4728	1.7816	0.0762	164	118
2005/2006	-1.4375	86.7735	33.4981	1.7189	0.0835	187	136
2006/2007	-2.9039	87.2969	31.4728	2.0515	0.0453	129	118
2007/2008	-1.6157	90.8493	31.4993	2.1030	0.0580	149	113
2008/2009	-2.3462	89.0871	28.9791	2.2041	0.0640	178	110
2009/2010	-1.9381	89.1398	30.2111	2.1959	0.0547	181	98
2010/2011	-2.8711	87.1813	31.0315	2.2748	0.0866	145	90
2011/2012	-2.3150	89.3447	31.1831	2.3016	0.0844	139	86

Table 1.2: Sample means for the meteorological variables grouping the days by the hydrological year (from October 1st to September 31st following year). The last two columns are the number of days with discharge and the number of days without any precipitation, respectively.

from October 1st to September 31st of the following year. These are shown in Table 1.2, the means of the year 2001/2002 do not appear in the table because the data starts at January 21st, 2002. Apparently, there is a positive trend in the mean daily discharge since 2006/07. Also, the daily accumulated precipitation shows an important increase along the years. It seems that there is not trend in temperatures. Regarding the number of days with discharge, it seems that there is no trend, however the number of days without precipitation decreases since 2006.

Fig. 1.5 shows the boxplots of the average daily mean glacier discharge divided by 11-day groups for a smoother description. Regard that there is a clear seasonality pattern. First, we

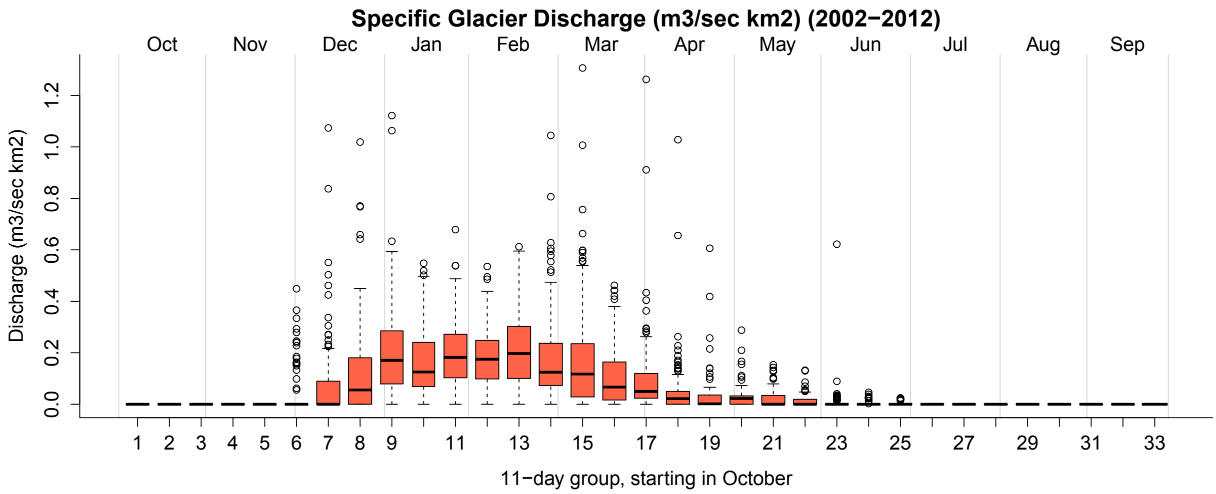

Figure 1.5: Boxplots of the average daily glacier discharge from 2002 to 2012 divided in 11-day groups.

observe that during the austral winter, which starts at the end of June, there is almost no glacier discharge. This produces the previously mentioned large amount of zero values of the discharge observations. The period of positive discharge begins at the end of the austral spring, between November and December. The extreme values which appear during this initial discharge period are usually known as "spring events" or "burst" (Warburton and Fenn (1994)), these are brief and violent episodes produced when the glacier brutally releases a large amount of water. They generally disappear in a few hours. They are explained by the plasticity of ice: when the previous year discharge wave finishes and the inner conduits of the glacier stop draining water; they tend to close by deformation, plugging the horizontal sections of the underground drainage network. On the other hand, this plugging does not take place in vertical shafts and, by slow percolation, they can get to fill during the Austral Winter, accumulating a latent hydraulic load that is freed at the beginning of the following Summer. When the sun rises at the beginning the following Austral Summer, solar radiation favours the glacier sliding, increasing it. The bottom seal of the vertical wells, loaded with water, is broken, and a violent discharge of their content is produced. Fig. 1.5 also shows that the maximum values for the daily discharge are observed during the austral summer, from the end of December to the middle of March. The daily discharge starts to decrease

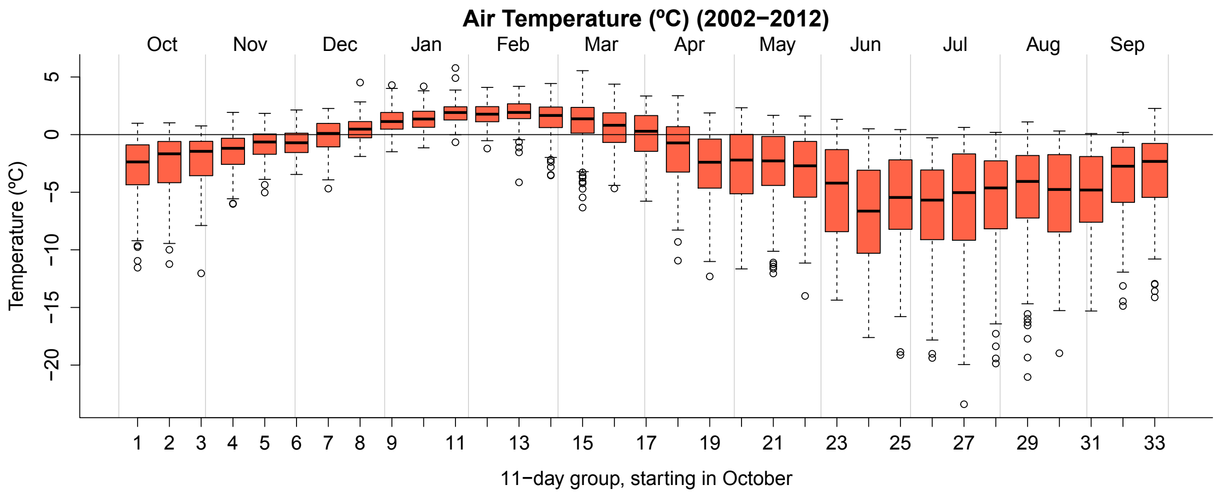

Figure 1.6: Boxplots of the average daily temperature from 2002 to 2012 divided in 11-day groups.

with the arrival of the autumn, at the end of March. However, we may also observe extreme values during this period, which are known as "aftershocks" (Warburton and Fenn (1994)). These are peaks in the glacier drainage that may appear when the discharge seems to be over and are typically caused by alternation of cold and heat episodes that cause fluctuations in the habit of the hydric discharge.

Fig. 1.6 shows the boxplots of the average daily temperatures divided, again, for a smoother description, by 11-day groups (Whitfield et al. (2002)). As before, we can observe a clear seasonality effect. Note that the average temperatures are above zero only during the austral summer, from the middle of December to the middle of March. Also during the summer period, we can observe less dispersion and more symmetry than in the rest of the year. On the contrary, temperatures start to decrease with the beginning of autumn and their dispersion increases. They are almost always below zero during the austral winter, from the end of June to the end of September, when they present a strong left asymmetry. These plots have been obtained with the help of the R package `seas` (Toews et al. (2007)).

Fig. 1.5 and 1.6 also show that there is an apparently clear relationship between temperature and glacier discharge. Although we can observe that this dependence is apparently not constant through time. During the austral winter, when the temperatures are very low, there is no glacier discharge. However, as commented before, the period of positive discharge starts in spring, when the temperatures increase, and during the austral summer, the median discharge values reach their maximum values when the largest values for temperatures are registered. Therefore, it is clear that there is also a seasonal dynamic in this dependence. This is also illustrated in Fig. 1.7 where the scatterplot for the temperatures and discharges are shown separately for each season. We can observe that there is a strong dependence in summer that disappears in winter. We can also observe that this dependence seems not to be linear.

Fig. 1.8 shows the time evolution of the Kendall's tau over a rolling window of 270 days. Observe that clearly the dependence is not constant along time, but it evolves in time describing cycles for each hydrological year. First, note that the coefficient reach their minima values between July and August which correspond to the Austral winter. In addition, we can observe two maxima every hydrological year corresponding to the beginning of the discharge periods in November (spring-events) and to the end of the discharge periods between April and May (aftershocks). Finally, observe that between these maxima the dependence is larger, as we expected, given that it corresponds to the summer time.

Fig. 1.7 also shows the same scatter plots on copula scale. These are obtained using the empirical cumulative distribution function evaluated at the observed temperatures and discharges for each season. Observe that the support of the copula function does not cover the whole unit square in some seasons due to the zero values observed in the glacier discharge.

Fig. 1.9 shows the scatter plot of each pair of all the variables and the histogram of each individual one for those days when the discharge was positive. Apparently, there are strong relationships between the variables although these seem not to be linear. The lower panel shows the tau-rank correlations,whose size is proportional to its absolute value, between each pair of variables. Then, we suggest the use of copulas to model these non-linear relationships.

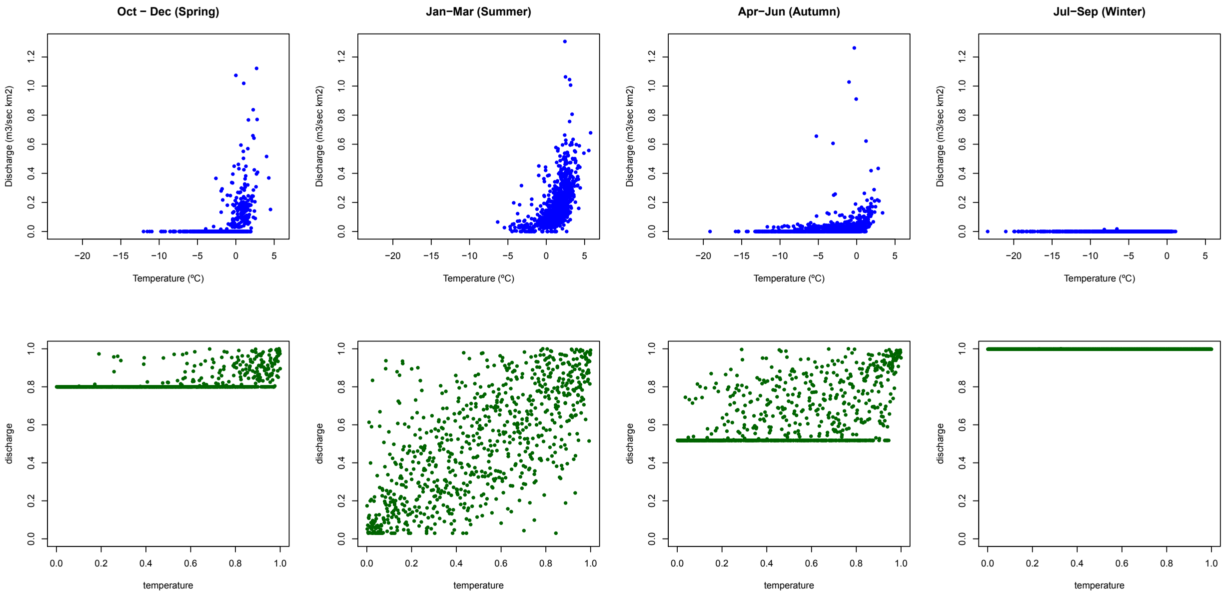

Figure 1.7: Scatter plots for the daily average temperatures and discharge in each season. The bottom line is for the scatter plots on copula scale with the use of the empirical distribution functions.

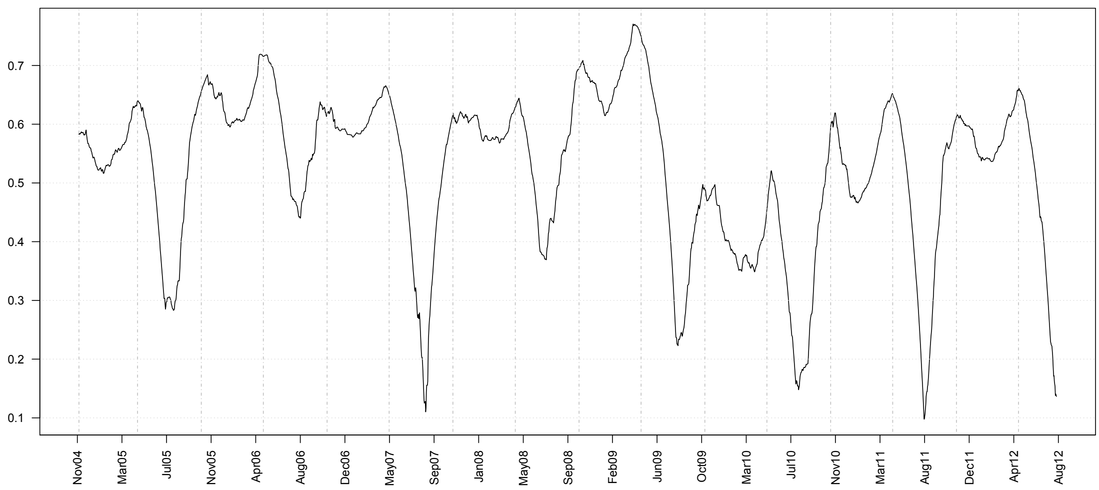

Figure 1.8: Kendall's tau of daily specific glacier discharge and average temperature. Taking a rolling window of 270 days.

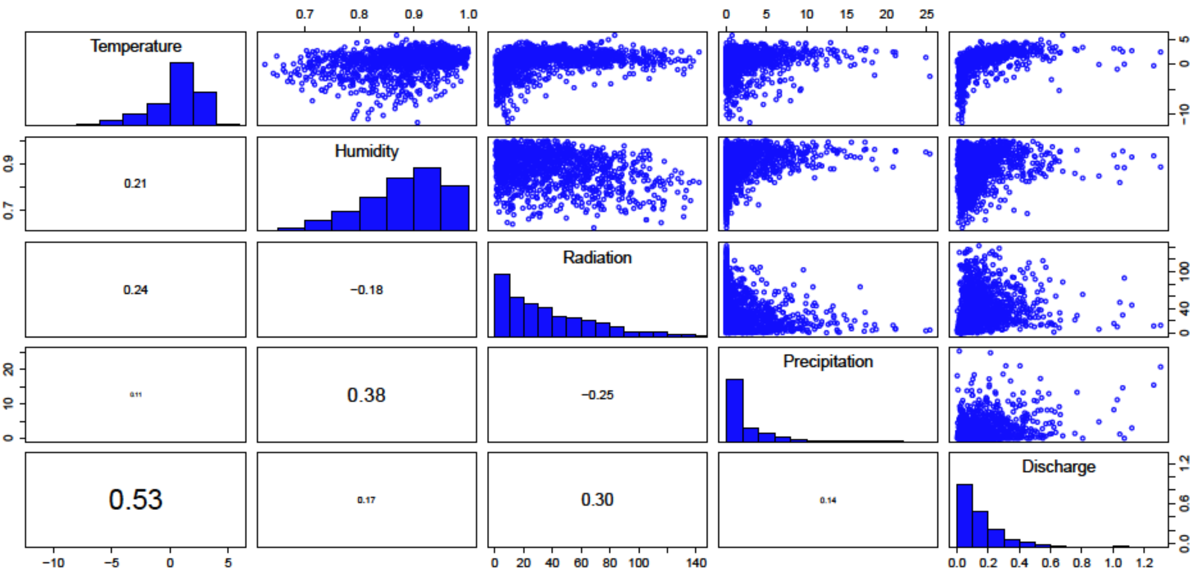

Figure 1.9: Scatter plot, histograms and Kendall's of the five variables when the values of the discharge are larger than zero. Size of the values of the is proportional to its absolute value.

1.3 Copulas

1.3.1 Introduction

Common characteristics of multivariate data include heavy tails, non linear dependencies and non stationarity, that suggest not to use Gaussian multivariate models. Copulas are statistical instruments that let us to model the relationship among the variables independently of the marginal distribution choice (Genest and Favre (2007)). In the last years, copulas have become a frequently used tool to model the non linear and non stationarity dependence, since they allow for modeling dependence among random variables in an easy conceptual way. Although the copula concept appears by the late 50s, the computer development in the last decades has caused a fast growth of the number of scientific papers related with copulas. In the last five years, more than 3000 papers related with copulas have been published with a constant year-on-year growth and, approximately, ten per cent of these publications are related with environmental sciences. See Nelsen (2007) and Romera and Molanes (2008) as extensive reviews about copulas.

Sklar (1959) used the term copula firstly in a French paper, followed by a similar English paper (Sklar (1973)) and comes from the Latin word *copulare* that means to joint or to connect. The basic idea of copulas is to separate the dependence structure among the variables from their marginal distribution functions. Alternatively, copulas can be viewed as a way to connect the dependence structure with the marginal distribution functions using the Sklar's theorem. Moreover, different copulas can be used to connect previously defined marginal distributions to define different multivariate distributions with distinct dependence structures.

Copulas have been widely used in many different fields. In climate and weather research (Schoelzel and Friederichs (2008); Cong and Brady (2012)), in hydrology (Genest and Favre (2007)). Recommended readings about this topic are Embrechts (2009), Genest and Favre (2007). Patton (2009) reviews the use of copulas in econometric models and Genest et al. (2009) provide an interesting bibliometric overview. Standard monograph is Joe (1997). Rosenberg and Schuermann (2006) show an approach based on copulas for management risk with heavy tailed asymmetry data.

Copulas provide a good tool to model the portfolio assets with better results than the approach based on cross correlation. Kole et al. (2007) analyzes the importance of selecting a precise copula and introduce an extension of the standard good-of-fit tests to copulas. In most of these works, copula models are static, that is, their parameters remain constant along time. Time-varying copulas have been widely used in finance (Patton (2012); Ausin and Lopes (2010); Manner and Reznikova (2012)) but, up to our knowledge, their use is very limited in hydrological research.

1.3.2 Copula definition and properties

Let be random variables with marginal distribution functions , and joint distribution then, Sklar's theorem proves that it exists a d-dimensional copula such that

(1.1)

Conversely, if is a -dimensional copula and are continuous distribution functions, then is a joint distribution function with marginal functions

A copula is a distribution function by definition, if this function is derivable, we can calculate the copula density as usual,

(1.2)

where are the marginal density functions of

Observe that the Sklar's theorem let us to define joint distribution functions in two steps. Firstly, we specify the marginal distributions of each variable and, secondly, we joint them through a copula function which models the dependence structure among the variables.

If the marginal distribution functions are continuous, it can be proved that the copula function is unique. Otherwise, there exists a copula function but it is not unique (Embrechts et al. (2005)).

1.3.3 Dependence measures

Different measures of dependence have been introduced in the literature to quantify the degree of relationship between two random variables. Examples are Pearson's linear correlation coefficient and the rank correlations of Kendall and Spearman. As commented previously, the copula of a multivariate distribution describes the dependence structure between its variables. Thus, it is very convenient to express the dependence in terms of the underlying copula, but this is only possible if the dependence measure is invariant under strictly increasing transformations. The Pearson linear correlation cannot be expressed as function of the underlying copula because it does not verify this condition.

Both, the coefficients of Kendall and Spearman, measure a kind of dependence called concordance. A pair of random variables is said to be concordant if large (small) values of one variable tend to be associated with large (small) values of the other. More formally, if and are two observations from a vector we will say that they are concordant if

Population Kendall's coefficient, , is defined as the difference between the probability of concordance and discordance of and , then we will have:

where and are independent vectors of random variables with joint distribution function

In the other hand, Spearman's coefficient, , is defined as the linear correlation between and , thus:

where, again,

Genest and Rivest (1993) introduced a method for measure the dependence, the -function. The -function is characteristic for each copula family and is defined as:

(1.3)

where is the Kendall's distribution function for the copula with parameter , and is distributed according to . Comparing empirical to theoretical -functions gives a method to select which copula family might be appropriate to describe the observed dependence. For Archimedean copulas, which will be defined in 1.3.4, there is a closed form expression in terms of the generator function of the copula,

where is the derivative of . For more details see Schepsmeier (2010). For the bivariate Gaussian and t-copula no closed form expression for the theoretical -function exists, although it can be simulated.

Another rank-based graphical tool for visualizing dependence is the K-plot. It shows the empirical information and, superimposed on the graph there is a straight line corresponding to the case of independence and a smooth curve associated with perfect positive dependence, the proximity of empirical points to that lines indicates the correlation between the points. More detailed information can be found in Genest and Favre (2007).

1.3.4 Families of Copulas

The most simple examples of copulas are those that model extreme cases of independence and perfect positive or negative dependence between the variables. The independence copula is given by:

where the joint distribution is the product of the marginal distribution functions of each variable. In the opposite side, we have perfect positive dependence, through the so called comonocity copula which is given by,

In the bivariate case we can have perfect negative dependence. The copula that models this situation is called countercomonocity copula and its expression is:

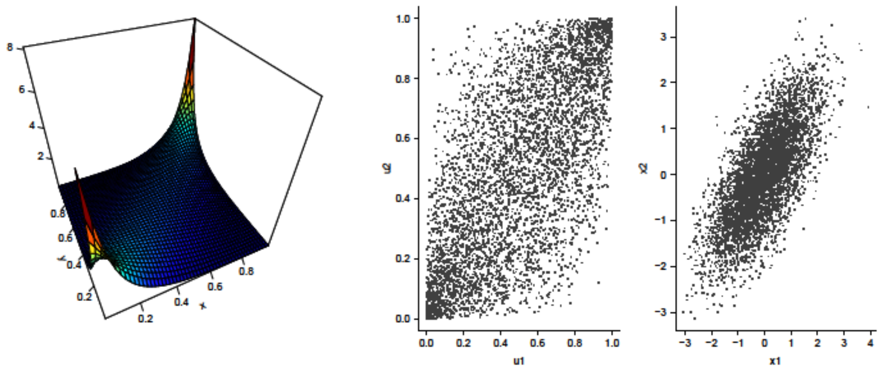

Figure 1.10: Density function (left), scatterplot of (middle) and scatterplot of (right) of a bivariate random variable whose marginal distributions are standard Gaussian and their dependence is modeled by a Gaussian copula with parameter , corresponding to

The most famous family of copulas is the class of elliptic of copulas, which use the relationship that exists in the multivariate elliptic distribution functions. In this family, we have the Gaussian copula, whose expression in the bivariate case is

(1.4)

where is the standard Gaussian distribution function, the standard Gaussian bivariate distribution function with correlation If the marginal functions are also normal, this joint bivariate distribution will be the bivariate Gaussian distribution. Fig. 1.10 shows the density function and scatter plot of a bivariate Gaussian copula. It also shows the scatter plot of a sample from a joint distribution with standard Gaussian marginal distributions and a Gaussian Copula, which obviously corresponds to a bivariate Gaussian distribution.

Another copula from the elliptic copula family is obtained with the bivariate t-Student structure, it is named the t-copula and can be expressed as

(1.5)

where is the bivariate t-Student distribution function with degrees of freedom and correlation , is the univariate distribution function of a t-Student distribution with degrees of freedom and correlation .

The parametric copulas most used belong to the family of archimedean copulas that capture a great variety of dependence structures. The archimedean copulas are built from a continuous strictly decreasing function named the copula generator function. The archimedean copula in terms of its generator function is given by,

In this kind of copulas, we have the Gumbel copula, whose expression is,

(1.6)

where . If this copula coincides with the independence copula and if , the Gumbel copula tends to the comonocity copula. One of the main advantages of the Gumbel copula is that it allows for right tail dependence (Embrechts et al. (2001)). Fig. 1.11 illustrates this copula. Another example of archimedean copula is the Clayton copula, that is defined by

and it is shown in Fig. 1.12. This copula is defined for but different form 0. If it coincides with the independence copula, and for it coincides with the countercomonocity copula. Another case is the Frank copula, which function is:

and which density function and scatter plot are drawn in Fig. 1.13, the parameter for this copula

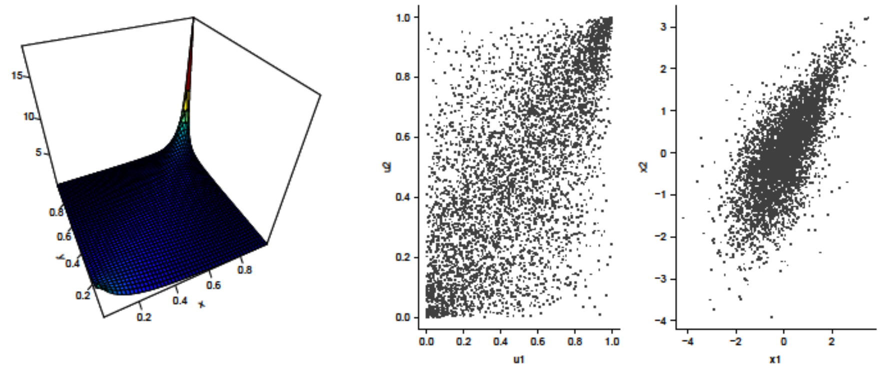

Figure 1.11: Density function (left), scatterplot of (middle) and scatterplot of (right) of a bivariate random variable whose marginal distributions are standard normal and its dependence is model by a Gumbel copula with dependence parameter , corresponding to .

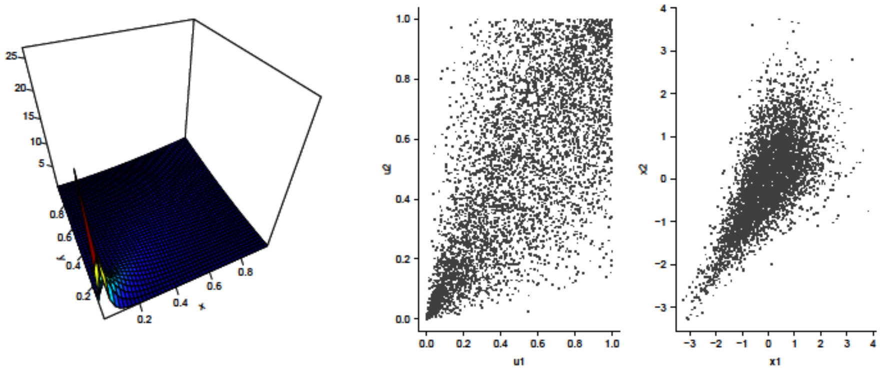

Figure 1.12: Density function (left), scatterplot of (middle) and scatterplot of (right) of a bivariate random variable whose marginal distributions are standard normal and its dependence is model by a Clayton copula with dependence parameter , corresponding to

can be any real number except zero, it has more symmetry than previous copulas and has peaks in right upper and left lower of , although they are less pronounced than in the previous copulas. Finally, we have the Joe copula, whose expression is

and is also defined or any real number different of zero.

The Archimedean copulas have rotated versions that capture the relationship between the

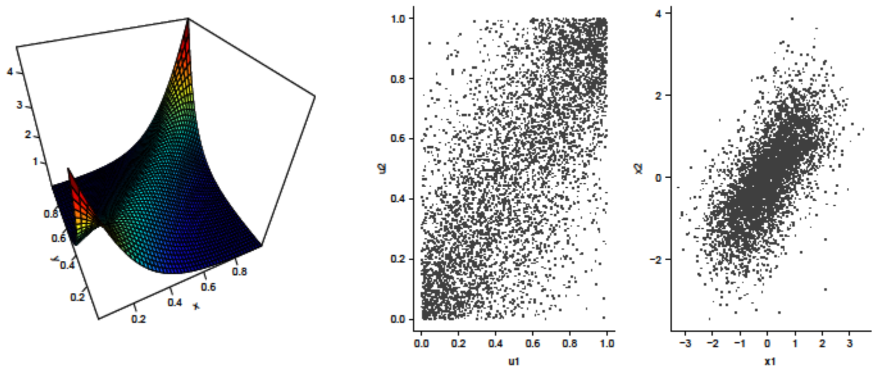

Figure 1.13: Density function (left), scatterplot of $u_i = \phi^{-1}(x_i)$ (middle) and scatterplot of x_i (right) of a bivariate random variable whose marginal distributions are standard normal and its dependence is model by a Frank copula with dependence parameter $\theta = 5.7363$, corresponding to $\tau = 0.5$.

Copula	Generator	Kendall$-\tau$
Gumbel	$\phi = (-\log(t))^{\theta}$	$\tau = 1 - \frac{1}{\theta}$
Clayton	$\phi = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$	$\tau = \frac{\theta}{\theta+2}$
Frank	$\phi = -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$\tau = 1 - \frac{4}{\theta}\left(1 - D_1(\theta)\right)$
Joe	$\phi = -\log\left(\log(1-(1-t)^{\theta}\right)$	$\tau = 1 + \frac{4}{\theta^2}\int_0^1 x \cdot \log(x) \cdot (1-x)^{\frac{2(1-\theta)}{\theta}} dx$
Gaussian		$\tau = \frac{2}{\pi}\arcsin(\theta)$

$D_1(\theta) = \frac{1}{\theta}\int_0^{\theta} \frac{x}{e^x - 1} dx$ is the so called Debye function.

Table 1.3: Archimedean copulas: Copula generator and relationship between de Kendall$-\tau$ and the copula parameter. Also includes the relationship between τ and θ in a Gaussian copula.

variables in the different corners of the unit square. The parameters of the archimedean copulas can be expressed in terms of the Kendall$-\tau$ as it is shown in Table 1.3, which also shows the generator function for each copula.

1.3.5 Vine copulas

Note that almost all the previously cited families of copulas are bivariate. Although there exist multivariate versions for some of the previously cited copulas, we prefer vine copulas because they are flexible structures that decompose any multivariate copula in a set of bivariate copulas, and they are useful to analyze the relationship among more than two variables. Joe (1997) was the first in suggesting this approximation, whereas Bedford and Cooke (2001) introduced a graphical structure, called regular vine structure, to help to organize the different pairs of copulas. More specifically, a vine copula for variables is a structure composed of trees, where the edges of one tree are the nodes of the next tree. Different bivariate copulas can be selected for each edge, introducing more flexibility. In particular, we will consider a family of vine copulas, known as canonical vines (c-vine), where in each tree there is always a node that is connected to all other. Aas et al. (2009) call pair copulas to the bivariate copulas and pair copula constructions (PPCs) to the vine copulas. See Haff et al. (2010) for a review.

Multivariate distributions can be decomposed recursively as a product of conditional distributions. If we have , a set of random variables with joint distribution and joint density function . We can consider the decomposition:

(1.7)

where is the conditional distribution function and is the conditional density function. Also, using Sklar's theorem (1.1) in dimension two, we can obtain,

(1.8)

where is the bivariate copula density function that models the dependance between and . Using the equation (1.8), we can express the conditional distribution function of given as,

(1.9)

Then, using the expression (1.9) in a recursive way, we may construct the conditional distribution of given :

where, to simplify notation, we have denoted,

Using this expression in (1.7) and taking and we have that:

According to Bedford and Cooke (2001), this structure is called canonical vine distribution or C-vine. For example, the construction of a C-vine copula for is

(1.10)

By the Sklar theorem (1.1) we know that

(1.11)

and

(1.12)

and

(1.13)

where

and replacing (1.11), (1.12) and (1.13) in (1.10), the four-dimensional joint density can therefore be represented in terms of density of bivariate copulas.

Bedford and Cooke (2001) saw that they could draw these pair-copula decompositions with a sequence of connected nodes that is called vine tree, where the edges between two nodes indicate the indices used for the copula marginal densities. Fig. 1.14 shows the graphical representation of a C-vine copula whose density function is given by:

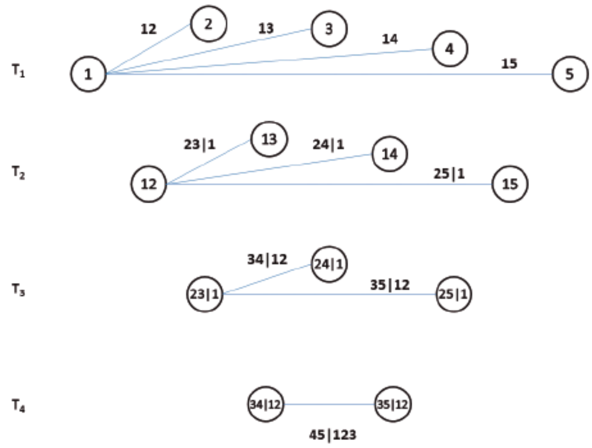

Figure 1.14: C-vine copula representation for the case of m=5

Vine copulas have been successfully used in a few number of papers in hydrology. For example, Gyasi-Agyei and Melching (2012) model the internal dependence structure between net storm event depth, maximum wet periods depth, and the total wet periods duration. Gyasi-Agyei (2013) models the dependence between total depth, total duration of wet periods, and the maximum proportional depth of a wet period in a rainfall disaggregation model. Xiong et al. (2015) study the dependence between annual maxima daily discharge, annual maxima 3-day flood volume and annual maxima 15-day flood volume to understand the change-point detection of multivariate hydrological series. Note that these papers deal only with three hydrological variables.

1.3.6 Copula Classical Inference

Parametric models

Assume that we observe data, , from a multivariate distribution with marginal density function , marginal distribution function for , and copula . Then using Sklar's theorem (1.1) and the joint density function (1.2), we can get the log-likelihood, that can be written as,

(1.14)

where

is the log-likelihood contribution to the dependence structure of copula , while

is the log-likelihood contribution of each marginal. Note that under the assumption of independence, the log-likelihood would be .

The MLE is obtained by maximization of the function (1.14), but this maximization can be very complicated especially if we have too many parameters. Joe (1997) gives an alternative to the simultaneous maximum likelihood of marginal functions and copula parameters by splitting the process into two steps. The approach is given by the decomposition . This method is called Inference Functions for Margins (IFM). In the first step, the marginal parameters are separately estimated by maximum likelihood, as if the variables would be independent. Then, in the second step, we estimate the copula parameters through the maximization of , plugging-in the estimated marginal parameters by the ones we have estimated in the first step, that is:

1. Obtain the MLE parameters of the marginal distributions

   for

2. Obtain the MLE parameters of the copula

Joe (1997) shows that MLE and IFM methods are equivalents in the case that the marginal functions are Gaussian and the copula is the Gaussian copula. Serfling (2009) shows that IFM estimator is consistent and asymptotically normal under the usual conditions or regularity in the multivariate model and in their marginal functions.

Semiparametric models

They are based, again, in the decomposition First, we estimate the marginal functions using non-parametric techniques. Next we estimate the copula parameters by the maximization of the log-likelihood of the dependance , that is, we need to maximize

where are the non-parametric estimators of the marginal functions

Genest and Rivest (1993) demonstrate that this semiparametric estimator, , is consistent and asymptotically normal under the appropriate regularity conditions. Besides, they suggest a consistent estimator for the variance-covariance matrix of .

Non-parametric inference and empirical copula process

The most popular non-parametric procedures for copulas are based on the inversion formula that appears in Durante and Sempi (2010):

where is the pseudo-inverse of Therefore an estimator of the d-copula will be:

where is a non-parametric estimator of the distribution function . Typically is taken as the empirical distribution function

—

and the are estimated using the pseudo-inverse functions of the univariate empirical distributions – or their reescalated versions.

Deheuvels (1979) and Deheuvels (1981) establish the consistence and the asymptotic Gaussianity on the empirical copula process for random vectors observations with independent marginal functions. Gaenssler et al. (2013) and Scaillet and Fermanian (2002) establish the consistence and asymptotical normality of general copula process with continuous partial derivatives. Scaillet and Fermanian (2002) also show that, under regular conditions, the asymptotical normality is verified for smoothed copula process. Notice that the empirical copula is not a smooth function. One way to smooth this empirical copula can be performed using the Bernstein polynomials. These can approximate the distribution functions leading to the empirical Bernstein copula of Sancetta and Satchell (2004).

1.3.7 Bayesian inference for copulas

Parametric inference

When the data are continuous, the likelihood of n independent observations with distributed as is where

are the parameters of the marginal models. And

is the marginal density of . Most of the parametric copulas have analytical expressions from their densities that let us to get the estimation directly by maximum likelihood.

However, there are cases where the Bayesian approach is better: Primarily, if the marginal distribution or the copula functions are complex, it can be very difficult to maximize the likelihood directly. One solution can be to use the described IFM by Joe (1997). Another alternative method is to use an iterative scoring algorithm like suggest Song et al. (2005). However a bayesian alternative is to build the posterior joint distribution using Markov Chain Monte Carlo methods (MCMC), with the parameters and generated individually under a Gibbs Samplig schema, see Pitt et al. (2006), Dos Santos Silva and Lopes (2008) and Ausin and Lopes (2010).

Secondly, the hierarchical Bayesian approach gives good results in modelling multivariate data. For example, Pitt et al. (2006) extend the bayesian selection to cases with non-linear dependance. And Smith et al. (2010) use the hierarchical models for the marginal functions and for the joint estimation where the copula function captures the dependence structure.

Finally, when we are estimating a copula model the objective is often to make inference over the dependence measures, the quantiles and functionals of the random variable or its parameters . The evaluation of the posterior distribution of these elements can be obtained directly by MCMC methods.

As an alternative method to the MCMC, we have the Approximate Bayesian Computation

(ABC) algorithms. The name ABC appeared in Beaumont et al. (2002), but this approximation was firstly developed in Tavaré et al. (1997) and Pritchard et al. (1999). See Beaumont (2010) as an extensive review about ABC. This algorithm is based in the replacement of the likelihood function by the simulation of values from the model with random parameters and the comparison of these simulations with the observed data. If the distance between observed and simulated data is less than a small value, , we choose the random parameters as values of the posterior distribution of the model parameters. The distance between the sets of data is very often greater than , what produces a big amount of rejects. However, the use of summary statistics to compare the observed and the simulated data improves the method (Tavaré et al. (1997)). Additionally, Beaumont et al. (2002) propose a method based on regression for reconstructing the posterior sample of the parameters. The regression is developed only with the candidate parameters whose simulations provide the closest summary statistics. The candidate parameters are the dependent variables and the observed data summary statistics are the covariates. They also suggest to give more weight to the parameters whose distance between the simulation summary statistics and the ones of the observed data are smaller.

Turner and Van Zandt (2012) propose three different methods to perform the ABC: Rejection sampling, MCMC-sampling and Particle filter, taking into account the distance, from the simulation summary statistics to the empirical summary statistics, instead of the likelihood. Turner and Van Zandt (2014) propose a Gibbs sampling method as ABC method. This proposed model uses the distance between the summary statistics, smoothed by a kernel, instead of the posterior probability of the parameters to obtain the rate acceptance. As in the case of the classical Gibbs sampling, they also need a tuning parameter to control the rate of acceptance.

Non-parametric inference

According to Burda and Prokhorov (2014), the flexibility and feasibility of non-parametric Bayesian models based on infinite kernel densities have reached high popularity even in complex modeling scenarios. However, these models have been rarely used in more than one dimension because, in the multivariate case, there is a large number of parameters to estimate to establish the dependence among the variables. In their paper, they propose a factorization scheme of multivariate dependence structures based on the copula modeling framework, whereby each marginal dimension in the mixing parameter space is modeled separately and the marginal functions are then linked by a nonparametric random copula function. In particular, they use one-dimensional Gaussian mixtures for the marginal functions and multivariate Bernstein polynomial as a link function, under a prior Dirichlet process. They demonstrate that this scheme suppose an improvement on the accuracy of the estimated density with respect to the obtained using a Gaussian mixture.

Alternatively, Wu et al. (2014) show a non-parametric Bayesian approach of high dimension copulas. First, they introduced the skew-normal copula, that is later extended to a infinite mixture model. The skew-normal copula imposes some limitations over the Gaussian copula. Their approach is only for the copula function model. It can be considered as a non-parametric bayesian approach to the Genest and Rivest (1993) ideas.

1.4 Overview of thesis

In this thesis, we focus on copulas based models to show the non-linear and seasonal relationships among different meteorological variables such as air temperature, percentage of humidity, solar radiation, precipitation and glacier discharge. The main purpose is to obtain their joint distribution function. Then, we get the conditional distribution function of the discharge given the values of the other meteorological variables. The discharge conditional distribution will allow us to obtain predictive values of the glacier discharge.

Chapter 2 focuses on temperature and discharge and explores their seasonality, not only in their location, scale or shape but also in their relationship. A parametric copula model is proposed for the joint distribution. The Bayesian point of view has been performed to make inference over the model parameters, where all the parameters are estimated in a one single step, in contrast with the usual two-step approach. The contents of this Chapter resulted into a paper by Gómez et al. (2017), which has been published in *Stochastic Environmental Research and Risk Assessment*.

Chapter 3 includes into the model the variables of humidity, solar radiation and precipitation in addition to the temperature and the glacier discharge. In this model, the seasonality is captured dividing the data in different periods. Moreover, data in each period is split into different groups to overcome the problem with zero values in discharge and precipitation variables. The parameters are calculated for each period and group separately. Structures based on c-vine copulas have been selected to model the multivariate relationship among these variables. In this Chapter, classical inference, based on mle, is preferred to make inference over the model parameters. The contents of this Chapter resulted into a working paper by Gómez et al. (2016).

Finally, Chapter 4 supposes that the relationships between the variables are not independent in each period and each group. Then, a hierarchical model is proposed, where the relationship between each pair of variables is led by common hyperparameter, independently of the period or group they belong. Again, the Bayesian point of view is used to make inference over the parameter of this model. The ABC technics have been selected in this occasion.

Chapter 2

Seasonal copula models for the analysis of glacier discharge

In this Chapter, we present a general model to describe the joint seasonal dynamics for the temperature and the glacier discharge. Firstly, we define separately the marginal models for the temperature and the glacier discharge using time-varying periodic distributions. Then, we describe the seasonal dependence using a time-varying copula model whose parameters vary periodically along time.

2.1 Proposed model

2.1.1 Marginal distributions

Fig. 1.6 shows the yearly seasonality of the daily temperature at each day , which will be denoted by . In order to approximate this seasonal behavior, we assume that the distribution of changes periodically through time with a location parameter, , given by:

(2.1)

where is the annual periodic cycle. Observe that this is an approximation by a partial sum of a trigonometric Fourier series with terms, where the fundamental frequency is , the amplitude parameters are , where , and the phase angle parameters are , where . Note that each angle phase, , is only defined in the semi-unit circle since:

Fig. 1.6 shows that not only the mean of temperature varies periodically along time, but also the variance and possibly, the shape of the distribution. Therefore, we can assume that the parameters of of scale, , and shape, , also vary periodically along time such that,

(2.2)

where

are the amplitude parameters for the scale and shape, respectively, where and and is the same vector of phase parameters defined in (2.1) for the time-varying location. Note that it makes sense to assume that the phase vector is the same for the location, shape and scale, since we expect the same dynamics for the three parameters such that, for example, when the location increases, the scale and shape decreases. Note also that in (2.2), we have modelled the logarithm of the scale parameter, , to avoid that it takes negative values. Therefore, the set of parameters for the marginal distribution function of the temperature is given by and the number of Fourier terms, .

Once we have defined the periodic pattern for the location, scale and shape parameters, it is necessary to specify a distribution model for the time-varying temperature, . For example, we may assume a skewed normal distribution, (Azzalini (1985)), whose density is given by,

(2.3)

where and denote the pdf and cdf of a standard Gaussian distribution. Note that when , we obtain the symmetric Gaussian model, .

Alternatively, we can consider a generalized extreme value distribution model for the temperature, (Embrechts et al. (2013)), whose density is given by,

(2.4)

for when and for when . This is a very flexible distribution which includes the Weibull ( ), the Gumbel ( ) and the Frechet ( ) distributions as particular cases.

There are many other possibilities that could be considered to model the temperature distribution. In Section 2.2, we explain how to undertake model selection for the distribution model and for the number of Fourier terms from a Bayesian perspective.

Fig. 1.5 shows that the daily discharge also has a seasonal behavior. Then, we define a periodic time series model for the average daily discharge at each day , which will be denoted by . As before, we approximate the seasonal dynamics for the location, , and scale, , parameters of using partial sums of Fourier series:

(2.5)

(2.6)

where

are the amplitude parameters for the location and scale parameters and ,
, is the vector of phase parameters. Thus, the vector of parameters for the glacier discharge is given by and the number of Fourier terms, .

Clearly, we could also define a similar periodic dynamic for the shape parameter. However, for simplicity, we will only consider positive random variables with two parameters to model the glacier discharge. For example, we may assume a Log-Normal distribution for the glacier discharge whose density is given by,

(2.7)

Alternatively, we could assume a Gamma distribution, whose density is given by:

(2.8)

where the mean, and scale parameter are assumed to follow the seasonal dynamics given in (2.5) and (2.6), respectively. As commented before, model selection and parameter estimation will be addressed in Section 2.2.

2.1.2 Copula

As commented before and it is shown in Fig. 1.8, the dependence between the temperature and the glacier discharge is not steady along time. There is a strong dependence between this two variables in the austral summer and there is almost no dependence in the austral winter. In order to describe this pattern, in this Section we model the dependence between these two variables using a time-varying copula model. More specifically, we assume that the Kendall's tau coefficient, , follows a seasonal dynamic described by a periodic function given by,

$$ \text{——} \tag{2.9}$$

where , , are the amplitude parameters and , , are the phase parameters of the time-varying tau rank correlation parameter. Now the angle phase, , is defined in the unit circle since:

moreover, we impose the restrictions and to ensure that is always in the interval . This makes sense since the dependence between the temperature and the discharge will never be negative. Thus, the vector of parameters for the copula is given by and the number of Fourier terms, .

Different copula models could be used. For example, we might consider that the dependence structure is defined by a time-varying Gumbel copula, that is the Gumbel copula in (1.6) where its parameter vary over time. The copula is evaluated at and , where and are the marginal distribution functions for and , respectively, at time , as it is shown in Table 1.3, and the dynamics of are specified in (2.9). Similarly, we could consider many other parametric copula models with time-varying tau correlation, such as the Gaussian copula (1.4), that do not allow for tail dependence, where the copula parameter is . Another alternative would be to assume a Student-t copula (1.5), where the copula parameter is . However, this copula model impose symmetric tail dependence, which does not seem realistic in this context, and would also require to estimate the degrees of freedom as an additional parameter.

Therefore, assuming that the number of terms in each Fourier sum is known, the joint density function for the temperature and the glacier discharge at time will be given by,

(2.10)

where and represent the marginal density functions of the temperature and the glacier discharge, respectively, that can be specified for example using the distribution models given in (2.3) or (2.4) and (2.7) or (2.8), respectively, and where represents the copula density function whose corresponding cumulative distribution function can be specified for example using (1.6), (1.4) or (1.5).

2.2 Inference, prediction and model selection

2.2.1 Inference

Consider now the observed data series,

which provides the daily temperature and discharge measurements during days. Given these data, we would like to make inference on the model parameters, . In this Section, we first assume that the distribution models for the marginal distributions and the copula are known. Also the number of terms in the Fourier approximations, , and , are supposed to be known. Later, in Subsection (2.2.3), we will explain how to perform Bayesian model selection to select all of them, the distribution functions, the copula family and the number of Fourier terms.

If the data set were complete, the likelihood function would be just the product of the joint density functions, (2.10), for each . However, as commented in the data description, during the hydrological year 2003/04, it was not possible to register measurements for the glacier discharge since the external data-logger suffered flaws due to the hard meteorological conditions during the winter months. These values will be treated as missing data. In addition, there is a large amount of glacier discharge values that are recorded as zero. Considering that the glacier discharge is measured as a function of the river level, these zero values can be regarded as left-censored observations since they are actually smaller than a minimum value, , below which it is not possible to register any discharge value. We are assuming that the glacier discharge values in these cases are so small that they can not be registered accurately.

Therefore, the likelihood function for the model parameters is given by,

(2.11)

where represents a missing discharge values which are not available. The conditional probability for the glacier discharge can be obtained as,

where represents the partial derivative of the copula distribution function as described, for example, in Venter (2002),

Note that these correspond to the so-called h-functions defined in Aas et al. (2009). For example, for the particular case of a Gumbel copula, it is obtained that,

where is the Gumbel copula distribution function given in (1.6). And for the Gaussian copula, the function can be expressed as where denotes the Gaussian density function with mean and variance , and denotes the standard Gaussian density function.

In order to perform Bayesian inference, we must define prior distributions for the model parameters, . We impose proper but non informative prior distributions as follows. For each amplitude parameter, , we assume a large variance Gaussian prior , for , and . For each phase parameter, , we assume a uniform semicircular variable in , for and and uniform circular variable in , for and .

Given these priors and the likelihood specified in (2.11), it is not straightforward to derive analytically the posterior distribution, . Therefore, we use MCMC sampling strategies in order to obtain a sample from the joint posterior distribution of the parameters, which will allow us to develop Bayesian inference. We propose a Gibbs sampling schema which is carried out by cycling repeatedly through draws of each parameter conditional on the remaining parameters (Tierney (1994)). In particular, we select the Random Walk Metropolis Hastings (RWMH) algorithm for sampling from the conditional posterior distribution of the model parameters. We use a simple one-dimensional RWMH where each model parameter is updated separately using normal candidate distributions whose mean is given by the previous value of each parameter in the algorithm and whose variance can be calibrated to obtain good acceptance rates. The details of the proposed algorithm are explained in the Appendix A.

2.2.2 Prediction

Now, we are interested in estimating the predictive joint distribution of the temperature and discharge, , at any time . This can be done using Monte Carlo simulation based on the MCMC output. Consider a posterior sample of size of the model parameters, , for . Then, the values of the time-varying parameters, and , are known for each time and we can simulate values from as follows.

For each and .

1. Obtain the copula parameter from

2. Simulate a value from the copula:

3. Obtain the pair of values for the temperature and discharge:

Given this sample of the joint posterior distribution, we can obtain a sample from the marginal predictive distribution of the temperature by just taking the values . The posterior predictive mean and credible predictive intervals can be approximated using the sample mean for each and the corresponding 0.025 and 0.975 quantiles. Similarly, we can approximate the posterior predictive mean and predictive intervals for the glacier discharge.

Finally, we wish to estimate the conditional predictive distribution of the glacier discharge given a value for the temperature, , at any time . As before, this can be done by Monte Carlo approximation given the MCMC output as follows.

For each and ,

1. Obtain from the distribution selected for the temperature,

2. Find such that where

3. Set

Therefore, given a set of observed temperatures, , we can obtain a sample of the conditional predictive distribution of the discharge for each time point, . Using this sample, we can estimate the posterior predictive mean and credible predictive intervals for the conditional discharge using the sample means and the 0.025 and 0.975 quantiles of the sample as before.

2.2.3 Model selection

In order to compare different models, we use the Deviance Information Criterion (DIC). Models with smaller DIC should be preferred to models with larger DIC (Spiegelhalter et al. (2002)). This measure penalizes the effective number of parameters of the model. The DIC value is given by,

(2.12)

where the log-likelihood of the model parameters, , is given by:

Given an MCMC sample of size of the posterior distribution of the model parameters, , for , the DIC value, (2.12), can be approximated by,

2.3 Simulated data

In this Section, we exemplify the proposed methodology with one of the numerous artificial sets of data generated to examine our procedure. In order to simulate the data we have proposed a model, that is the number of components, the amplitudes and the phases for the parameters of the two variables and the Kendall between them. We have executed the algorithm with different configurations of the number of Fourier terms, assuming a Generalized Extreme Value distribution for the first variable, a Gamma distribution for the second one and a Gumbel copula

4	4	2	16353	1	4	2	17178
4	4	1	16479	1	4	1	17184
4	4	4	16953	1	3	3	17188
4	4	3	17005	1	3	1	17191
3	4	1	17044	1	3	2	17193
2	4	1	17140	1	1	1	17440
1	4	3	17162	1	2	1	18776

Table 2.1: DIC values for different number of Fourier terms, , and , assuming a GEV distribution for the temperature, a Gamma distribution for the discharge and Gumbel copula. For the simulated data.

for the relationship between them. Table 2.1 shows some of the results obtained. Note that the minimum value corresponds to terms for the first variable, terms for the second variable and terms for the parameter of the copula, that is, the configuration used to simulate the data. In addition, the algorithm has been performed assuming different models for the marginal distributions in both variables and also for the copula. For example, the minimum value assuming a Gaussian copula, with the GEV and Gamma as marginal distributions, is .

Once we have obtained that the best configuration is 4 components for both variables, 2 components for the copula, GEV distribution for the first variable, Gamma distribution for the second one and Gumbel as the copula, the MCMC for all the amplitudes and phases have been got. As an example, Table 2.2 shows the true values compared with the mean and credible intervals of these MCMC for the second variable. Fig. 2.1 shows the comparison between the estimated values for all the parameters and the true values used for the simulation of the data (left plot) and the boxplots of the the MCMC compared with the true values (right plots).

2.4 Results

In this Section, we illustrate the proposed methodology with the data provided by GLACKMA of measured discharge and temperature from October 1st 2002 to September 30th 2012. We have considered a large number of different models for the marginal and copula distributions that will

	True Value	Mean (int. cred.)
ψ_{12}	2.7	2.665 (2.609, 2.720)
ψ_{22}	2.8	2.793 (2.562, 3.025)
ψ_{32}	2.9	2.911 (2.672, 3.115)
ψ_{42}	3.0	0.879 (0.015, 3.084)
$a_{1\lambda}$	6.0	4.630 (3.171, 6.077)
$a_{2\lambda}$	3.0	2.089 (1.444, 2.921)
$a_{3\lambda}$	2.0	1.383 (1.015, 1.812)
$a_{4\lambda}$	0.5	-0.137 (-0.408, 0.401)
$a_{0\lambda}$	-6.0	-5.151 (-6.038,-4.184)
$a_{1\beta}$	-1.0	-0.355 (-1.827, 0.997)
$a_{2\beta}$	-2.0	-1.569 (-2.471,-0.939)
$a_{3\beta}$	-2.0	-1.622 (-2.093,-1.189)
$a_{4\beta}$	-0.5	0.203 (-0.544, 0.568)
$a_{0\beta}$	2.0	1.549 (0.605, 2.442)

Table 2.2: True values of simulated data and mean of the MCMC with the correspondent 95% credible interval for the second variable.

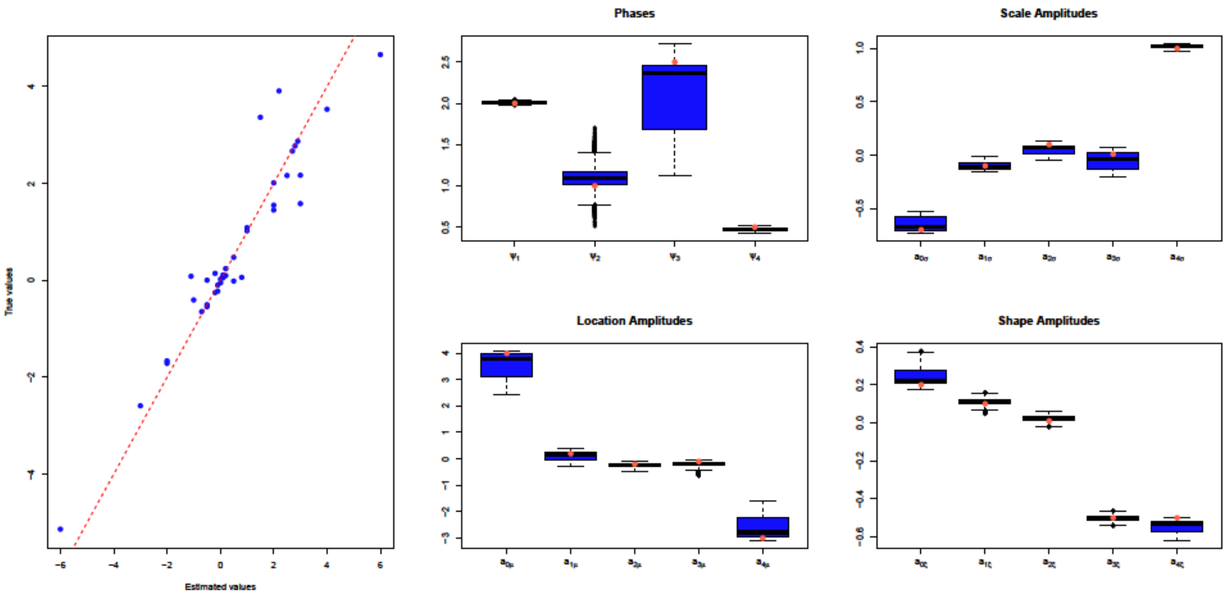

Figure 2.1: To the left, comparison between the parameters estimated by the model for the simulated data and the ones used to simulate these data. To the right, box plot of the MCMC of every parameter in the simulated data of the first variable, where red dots are the true values

Temperature		Discharge		Copula	
mean (sd)	int.cred.	mean (sd)	int.cred.	mean (sd)	int.cred.
2.016 (0.017)	(1.983,2.050)	2.731 (0.027)	(2.674,2.783)	2.269 (0.482)	(0.785,2.830)
2.943 (0.098)	(2.742,3.117)	2.841 (0.040)	(2.765,2.920)	1.588 (0.814)	(0.109,3.019)
2.865 (0.134)	(2.594,3.107)	2.783 (0.074)	(2.639,2.926)	0.821 (0.395)	(0.163,1.673)
0.646 (0.154)	(0.363,0.982)	2.916 (0.153)	(2.573,3.132)	-0.195 (0.216)	(-0.578,0.280)
4.002 (0.075)	(3.855,4.152)	6.692 (0.668)	(5.604,7.940)	-1.163 (0.291)	(-1.769,-0.628)
-0.265 (0.071)	(-0.407,-0.126)	3.378 (0.433)	(2.654,4.274)		
-0.325 (0.063)	(-0.456,-0.207)	1.638 (0.232)	(1.231,2.174)		
-0.152 (0.062)	(-0.274,-0.029)	0.447 (0.091)	(0.279,0.644)		
-2.777 (0.056)	(-2.886,-2.665)	-6.431 (0.406)	(-7.179,-5.749)		
-0.721 (0.018)	(-0.757,-0.687)	-0.889 (0.415)	(-1.653,-0.108)		
-0.136 (0.019)	(-0.173,-0.100)	-1.528 (0.225)	(-1.971,-1.093)		
0.102 (0.020)	(0.063,0.142)	-1.783 (0.168)	(-2.120,-1.459)		
0.014 (0.016)	(-0.018,0.045)	-0.432 (0.116)	(-0.665,-0.210)		
0.975 (0.013)	(0.949,1.002)	2.302 (0.273)	(1.786,2.795)		
0.261 (0.014)	(0.233,0.289)				
0.114 (0.016)	(0.081,0.145)				
0.004 (0.014)	(-0.025,0.032)				
-0.029 (0.011)	(-0.05,-0.008)				
-0.471 (0.009)	(-0.488,-0.454)				

Table 2.3: Model values for the parameters of the GEV distribution for the temperature, the Gamma distribution for the discharge and the Gumbel copula. Each parameter is obtained as the mean of its MCMC. The posterior deviation is the number between parenthesis. The third column of each parameter is the credible interval.

be discussed later. Firstly, we present the results for the preferred model according to the DIC criteria which consists of the GEV distribution described in (2.4), for the marginal distribution of the temperature with Fourier components, the Gamma distribution (2.8), for the marginal distribution of the discharge with Fourier components and Gumbel copula (1.6), with Fourier components.

The proposed MCMC algorithm is run for iterations, discarding the first as burn-in iterations. The chains have converged and they have good mixing. Table 2.3 shows the mean, posterior deviation and credible intervals for the model parameters.

Fig. 2.2 shows the observed discharge time series data, the posterior predictive means and credible intervals for the whole time period. Apparently the discharge is well modeled, for

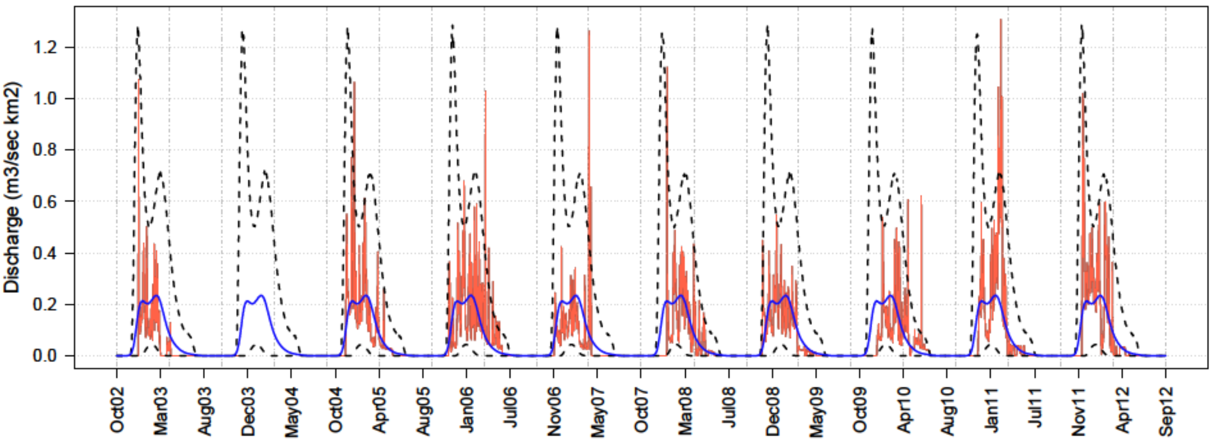

Figure 2.2: Observed discharge data, posterior predictive means and 95% credible intervals.

example the posterior means are very close to zero in those winter periods where no discharge is recorded and the length of the corresponding credible intervals are also quite close to zero. On the contrary, during the summer periods, the posterior discharge means are far from zero and the credible intervals are wider. Also, we can observe that the proposed model captures the Spring events and aftershocks at the beginning and the end of each period, respectively. Finally, observe that the proposed method is also able to produce Bayesian estimates and credible intervals for the missing period during the hydrological year 2003/2004.

Fig. 2.3 shows the observed temperature time series, the posterior predictive means and credible intervals for the whole time period. Observe that the model can capture the left-skewness and larger variability during the austral winter. In contrast, note that credible intervals are more symmetric and narrower during the summer periods.

Fig. 2.4 shows the posterior mean and credible intervals of the Kendall's tau together with the observed values for the temperature and discharge for the whole time period. This figure illustrates how the dependence varies over time, we can see how larger values of tau correspond to higher values of the temperature and discharge. Similarly, smaller values of tau correspond to lower temperatures and periods with no discharge.

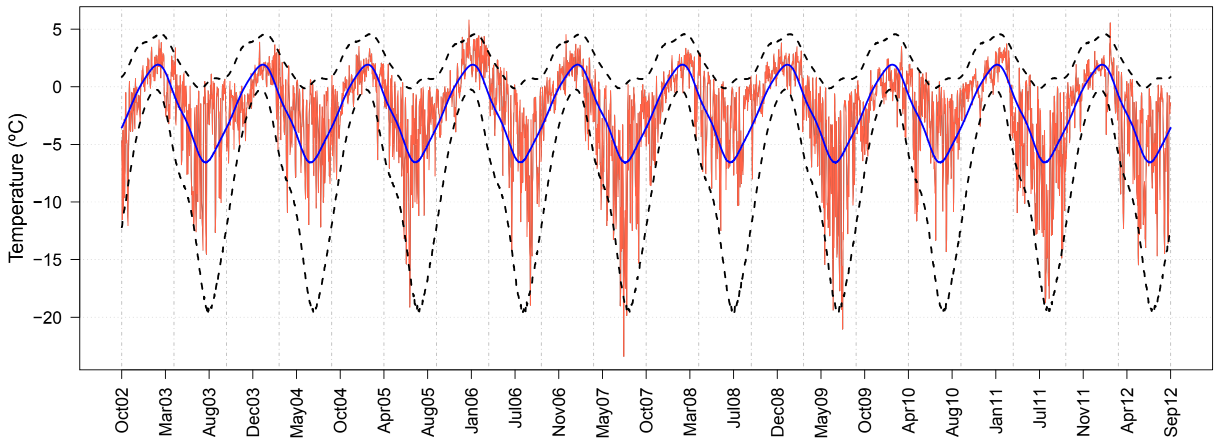

Figure 2.3: Observed temperature data, posterior predictive means and 95% credible intervals.

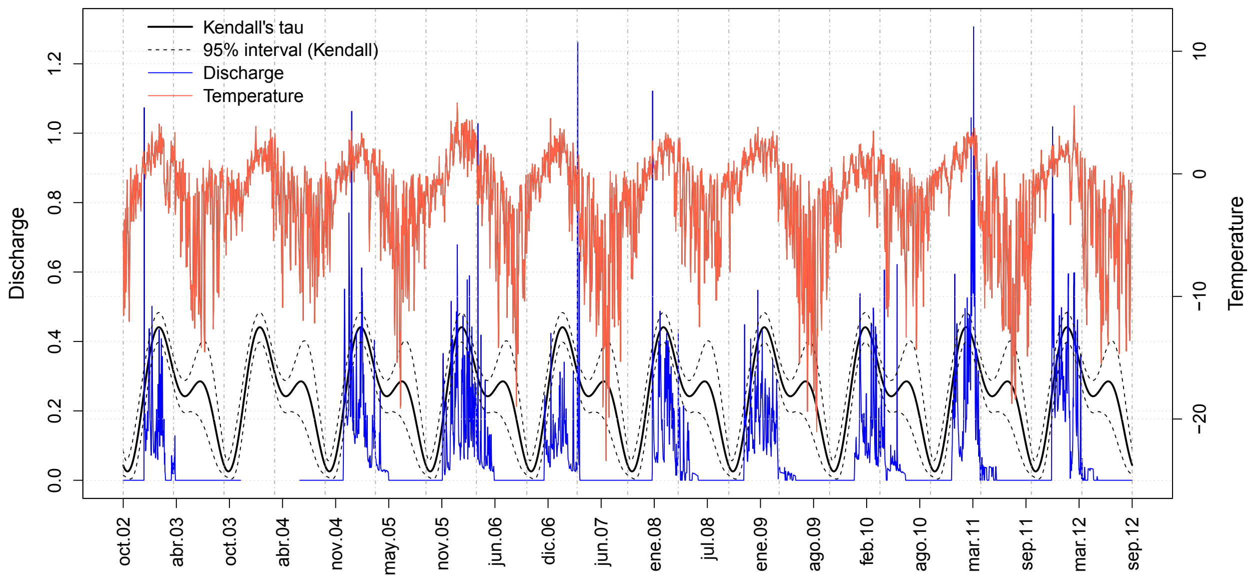

Figure 2.4: Posterior mean of the Kendall's tau and 95% credible intervals together with the observed values of the temperature and discharge for each time point.

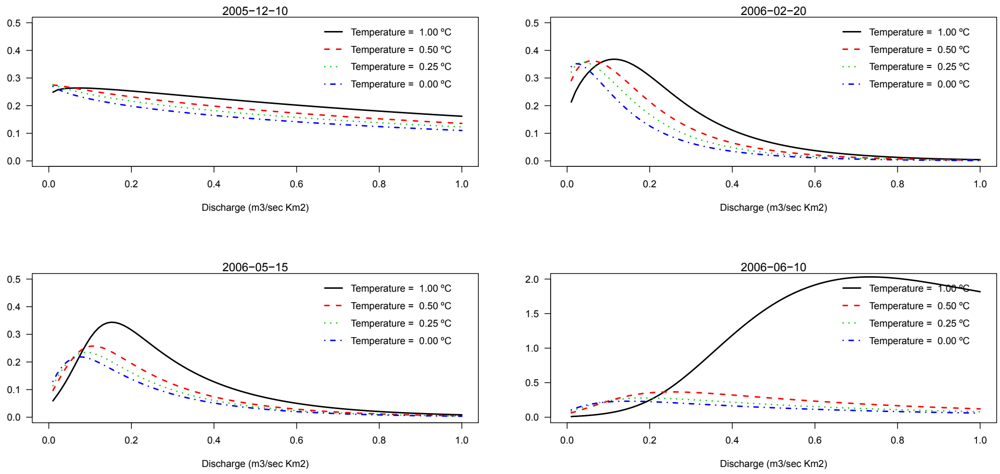

Figure 2.5: Conditional predictive density of the discharge given different values of the temperature for different days, at the beginning, in the middle and at the end of the discharge period

Now, we are interested in analyzing the influence of the temperature on the discharge. Observe that using our proposed approach, we can obtain estimations of the conditional predictive distribution of the discharge given any value of the temperature at any given time point. As an illustration, Fig. 2.5 shows the conditional density function of the discharge for different particular days, at the beginning, in the middle and at the end of the period of discharge given different values of the temperature. Note that, as expected, the larger is the temperature, the larger is the probability of observing large values for the glacier discharge, although the density plot is different depending on the day of the year. The density plot for the 10th of June and for one Celsius degree (in the last plot) is useful to explain the aftershocks, because that temperature is not very usual in that day.

Using the same approach, Fig. 2.6 shows the Bayesian estimations of the missing discharge values conditioned on the observed values for the temperature during the hydrological year 2003/2004 when the data-logger did not record the data appropriately.

Finally, observe that our proposed methodology also enables future predictions of both the joint distribution of discharge and temperature and the conditional discharge distribution given the temperature values. In order to illustrate this, Fig. 2.7 shows the estimations of the

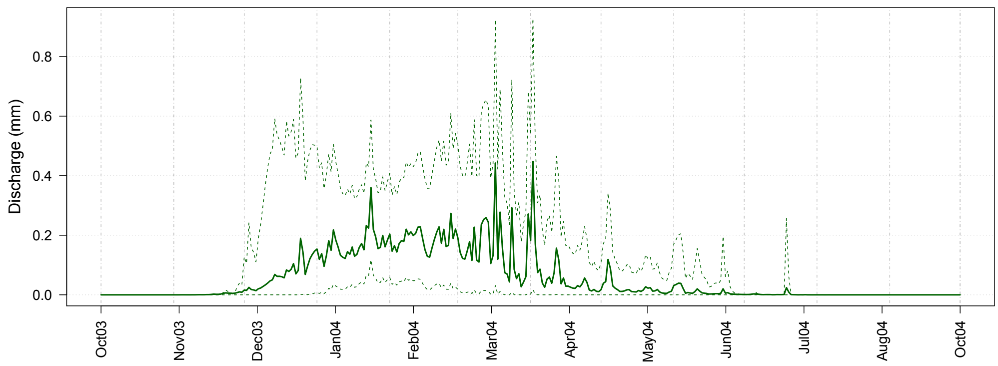

Figure 2.6: Predicted values for the missing discharge during the hydrological year 2003/2004 conditioned on the observed values for the temperature. Dotted lines represent the 95% credible intervals.

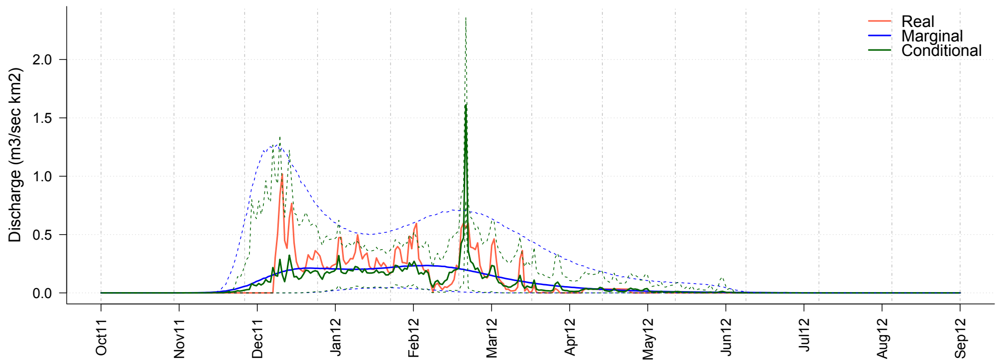

Figure 2.7: Observed data for the discharge, mean of the values of the predictive and 95% credible intervals. Hydrological year 2011-12.

predictive discharge distribution for the last hydrological year 2011/2012 given the information from previous years. These are compared with the true observed values during this year. Note that the predictive intervals always contain the true observed values. Fig. 2.7 also shows the estimations of the conditional predictive discharge during this last year given the values for the temperature. Observe that this provides in general better estimations for the discharge, although there is one single day where the temperature was extremely high which leads a large estimation for discharge.

4	4	2	14825	1	4	1	15043
4	4	1	14833	1	4	2	15044
4	4	4	14843	1	3	3	15115
4	4	3	14861	1	3	1	15135
3	4	1	14872	1	3	2	15178
2	4	1	14921	1	2	1	15259
1	4	3	15036	1	1	1	15750

Table 2.4: DIC values for different number of Fourier terms, , and , assuming a GEV distribution for the temperature, a Gamma distribution for the discharge and Gumbel copula.

2.4.1 Model selection

Finally, we illustrate how the model introduced before has been selected according to the DIC. Firstly, we put the emphasis on selecting the number of Fourier terms for the time-varying parameters of the temperature, , the discharge, and the copula, . Table 2.4 shows the DIC values for different choices of the number of Fourier terms assuming a Generalized Extreme Value distribution for the temperature, a Gamma distribution for the discharge and a Gumbel copula for the dependence. Note that the minimum value corresponds to terms for the temperature, terms for the discharge and terms for the time-varying parameter of the copula.

Similar tables have been obtained assuming different models for the marginal distributions of the temperature and the discharge and also for the copula. The number of selected Fourier terms is in general the same but the value of the DIC is larger in all cases. For example, the minimum DIC value assuming a Gaussian copula and the same marginal distribution models as before is . The same minimum DIC value is obtained for the t-copula since the estimated degrees of freedom are very large which implies that the obtained t-copula is very similar to the Gaussian copula. Finally, the is obtained for the model with the Clayton copula. Note that these values are larger than the minimum value obtained in Table 2.4 with a Gumbel copula which is given by , indicating that the Gumbel model is preferred than the other considered copulas.

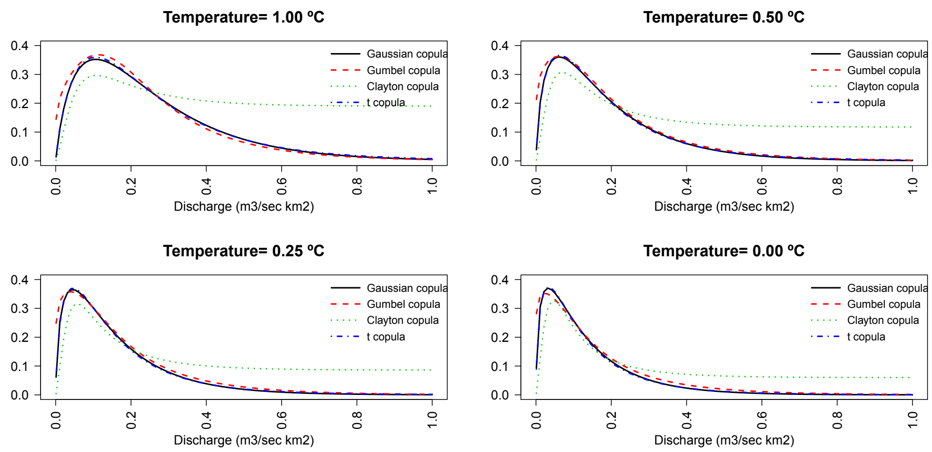

Figure 2.8: Conditional predictive density of the discharge for the models built with different copulas, given zero degrees as the value of the temperature for all of them and for one particular day in summer (02/20/2006).

In order to illustrate the differences among copula models, Fig. 2.8 shows the conditional predictive density of the discharge given different temperatures for one particular day in summer using the different copula models. This figure shows that the Clayton copula is not appropriate for these data, as expected, since this copula has not right tail dependence and it only allows for left tail dependence. On the other side, the obtained estimated models with the Gaussian and t-copula are very similar to that obtained with the Gumbel copula. However, it can be observed that the tail of the conditional distribution is slightly heavier with the Gumbel copula.

2.5 Conclusion and extensions

In this Chapter, we have proposed a seasonal dynamic model to describe the joint distribution of the glacier discharge and air temperature where not only the marginal distributions are time varying but also the relationship between these two variables is described by a time-varying copula. We have proposed a Bayesian procedure for making inference on the model parameters and prediction of the joint discharge and temperature distribution. Our approach allows for the simultaneous estimation of the marginal and copula parameters, which is in contrast with the classical two-stage estimation procedures.

An improved model could include structural changes over the time such that not only the model parameters were time-varying, but also the marginal and copula models could vary along time. For example, we could consider for each different season the possibility of using a different copula selections, Gumbel (1.6), Gaussian (1.4) or Student-t (1.5). Similarly, we could incorporate for different seasons the possibility of distinct marginal distribution models for the temperature and glacier discharge.

The proposed procedure could be extended to a multivariate model by including more environmental variables like precipitation, humidity or solar radiation. In this case, the use of multivariate copulas would be required. One possibility is the use of vine copulas.

The developed methodology could be also applied in other Pilot Experimental Watersheds installed by GLACKMA at different latitudes in both hemispheres, which could be compared with those obtained in this work.

Chapter 3

Vine copula models for the analysis of glacier discharge

In this Chapter, we introduce a method to predict future values of the glacier discharge given the observed values of other meteorological variables. First, we propose a copula model to describe the multivariate joint distribution of the five variables where, in addition, two of them have a large number of zero values. Then, we obtain the conditional probability of having no discharge. Finally, we derive the conditional distribution of the discharge given the other meteorological variables.

3.1 Proposed model

3.1.1 Multivariate copula model

Let and random variables, where is the temperature, the humidity, the radiation, the precipitation and the discharge. As commented in Section 1.2, in practice, it is usually observed that both, the precipitation and the discharge, have a large number of zero values. This fact has a quite important impact in the construction of our proposed model. Erhardt and Czado (2012) and Brechmann et al. (2014) propose a model for a multivariate

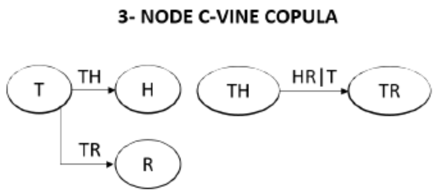

Figure 3.1: Structure of c-vine copulas with 3 nodes.

distribution in which the mixed variables are decomposed in zero inflated and continuous positive components. Following this idea, we define the joint distribution as a mixture of four different joint distributions, depending on the joint probability of presence of zero or positive values of the discharge and the precipitation. Thus, the joint density function of the multivariate variable is decomposed as,

with (3.1a)

with (3.1b)

with (3.1c)

with (3.1d)

Then, we define each of these four joint density functions in terms of copulas using the theorem of Sklar (1959). For example, (3.1a) can be expressed as,

(3.2)

where is the multivariate copula density describing the dependence structure in the variable and

(3.3)

and analogously for , , and . Where the superscripts denote that the variables are conditioned on zero discharge and zero precipitation.

Our proposal is to use a vine copula structure, defined in Subsection 1.3.5, for the multivariate copula. For example, the expression (3.2) can be decomposed, using the c-vine structure shown in Fig. 3.1, as,

where , and are the density functions of the bivariate copulas in each edge and

which are the conditional distribution functions of the uniform variable introduced in (3.3).

Similarly, the expression (3.1b) can be expressed in terms of copulas as,

(3.4)

where is the multivariate copula density describing the dependence of the variable and

and analogously for , , and , , , where the superscripts denotes that the variables are conditioned on positive discharge and zero precipitation.

The multivariate copula in (3.4) can be decomposed as a product of bivariate copulas, using the vine structure shown in Fig. 3.2, as,

(3.5)

where , , , , and are the density functions of the bivariate copulas in each edge,

and analogously for , and . Similar expressions can be obtained for (3.1c) and (3.1d) which are shown in Appendix B.

3.1.2 Marginal distributions

We now define a marginal distribution model for each one of the five meteorological variables , , , and . We decompose each variable in four cases according to the presence or not of precipitation and discharge, as in (3.1). For example, the density function of the temperature can be expressed as:

with

with

with

with

Then, for each of these four cases, we assume a parametric model based on finite mixture models. In particular, for the temperature, we consider finite mixture of Gaussian distributions. For example, the first density function can be written as,

and similar Gaussian mixtures for , and . Then, we may obtain , , and using the cumulative distribution function of a Gaussian mixture. Therefore, note that we have a set of parameters to estimate for each of the four Gaussian mixtures.

The same procedure is followed for the humidity and the radiation. Each of these variables is divided in four cases according to the presence or not of discharge and precipitation. Then, a finite mixture of Beta densities is selected for each of the four cases in the humidity and a finite mixture of Gamma densities for each case in the radiation. Their respective densities are,

and similar Beta mixtures for , and and Gamma mixtures for , and . Then, we may obtain , , , , , , and using the cumulative distribution function of a Beta mixture and a Gamma mixture respectively. Again, note that we have a set of parameters to estimate for each of the four Beta mixtures and a set of parameters for each of the four Gamma mixtures.

For the precipitation, given that it is greater than zero, two cases are differentiated corresponding with the presence or absence of discharge. Then, a finite mixture of Gamma densities is selected for each case, where the density function can be expressed as,

and a similar Gamma mixture for . Again, we may obtain and , using the cumulative distribution function of a Gamma mixture. Here, we only have a set of parameters for each of the two Gamma mixtures.

Finally, for the discharge and given that it is greater than zero, two cases are distinguished corresponding with the presence or absence of precipitation. Then, finite mixtures of Gamma densities are considered for each case whose density can be written as,

and similar Gamma mixture for . Then, we may obtain and , using the cumulative distribution function of a Gamma mixture. Here, we have a set of parameters for each of the two gamma mixtures.

Finally, note that the number of terms in each mixture, , may be different in each group and variable and they will be selected using a model selection criteria explained in Section 3.1.4

3.1.3 Conditional probability

Once we have defined the multivariate model given by the copulas and the marginal distributions, we may obtain many quantities of interest. For example, we may obtain the conditional probability of zero discharge for one particular day whose meteorological variables have been observed. Using the Bayes theorem, this probability is given by,

(3.6)

For the case when the precipitation is zero, the numerator in (3.6) can be expressed as,

(3.7)

where the first term is obtained from (3.1a). And the denominator of (3.6) can be expressed as,

where the first term is obtained from (3.7) and the second term can be obtained from (3.5) since it

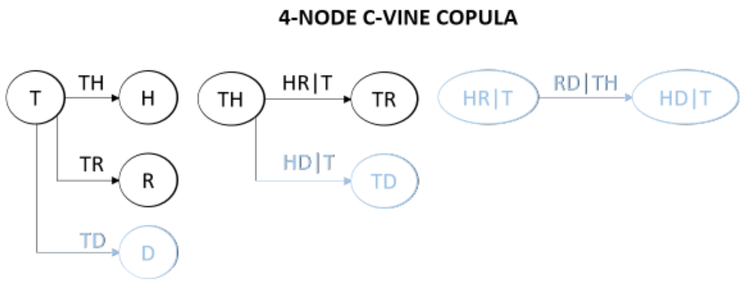

Figure 3.2: Structure of c-vine copulas 3 nodes inherited from a 4-node c-vine copula.

can be expressed in terms of a vine copula as,

and the terms of this expression already appear in the equation (3.5). Fig. 3.2 shows how the vine-copula for describing the dependence of the variable can be used to obtain the marginal dependence of .

Similarly, for the case when the precipitation is positive, the numerator of (3.6) can be expressed as,

(3.8)

where the first term is obtained from (3.1c). And the denominator in (3.6) can be expressed as,

where the first term is obtained from (3.8) and the second term can be obtained from (B.2) since it can be expressed in terms of a vine copula as,

(3.9)

As in the previous case, we can use the vine structure selected for describing the dependence of the variable to obtain the vine copula for describing the marginal dependence of the variable . Thus, all the terms in (3.9) can be found in (B.2).

Furthermore, using the defined multivariate distribution in (3.1), we may obtain the conditional distribution function of the discharge for given values of the meteorological variables, using the conditional probability in (3.6) as,

with

with

(3.10)

For the case when , the second part can be expressed in terms of the c-vine copulas as,

(3.11)

For the case when , the second part of (3.10) can be expressed as,

(3.12)

where,

and similarly for .

3.1.4 Parameter estimation and model selection

Now, given the set of data on discharge and other meteorological variables, we want to estimate the parameters for our proposed model (3.1). First, we divide the sample in four groups according to the presence or not of discharge and/or precipitation and estimate the probabilities of each group, which correspond to the joint probabilities of having or not zero discharge and/or precipitation, using empirical frequencies. The parameters can be different in each group, but the estimation procedure is the same. First, we select the number of mixture components of the marginal functions for each variable using the Bayesian Information Criterion (BIC), which generally penalizes number of parameters in the model. Then, these mixture parameters are estimated by the maximum likelihood method, separately for each variable. Next, the values of , for ; ; and , are obtained as explained in Subsection 3.1.2.

The next step is to fit the vine copula model to the data set. Note that here we assume the so called *simplifying assumption*, which imposes that each pair of copulas of conditional distributions does not depend on the values of the variables which are conditioned on. Although, this assumption has been criticized (Acar et al. (2012); Spanhel and Kurz (2015)), Killiches et al. (2017) propose the use of this assumption especially when the number of parameters is large. Moreover, Haff et al. (2010) came to the conclusion that vine copulas built with this assumption are "a rather good solution, even when the simplifying assumption is far from being fulfilled by the actual model".

In order to select a c-vine copula structure, we firstly need to set an order for the variables. As suggested by Aas et al. (2009), we try to set the variables with strongest dependencies in the first nodes of the tree. Therefore, we order the variables regarding on the values of the Kendall's tau. Thus, we have considered the temperature as the root variable, followed by the humidity, the radiation, the precipitation and, finally, the discharge. Appropriate pair-copula families are selected and estimated sequentially, using the BIC to determine the best copula family. For simplicity in the notation, we refer by to the different possibilities , for ; ; and , for example denotes the for the different cases of the temperature. The value of the parameters is estimated by maximum likelihood as follows,

1. Fit bivariate copulas for and , for , for all the edges in the first tree.

2. Generate the series , for , using the fitted copula from the previous step.

3. Fit bivariate copulas for and , for for all the edges in the second tree.

4. Using the same procedure, generate series from the edges and fit copulas between the nodes for the remaining trees.

The copula for each node can be selected between an elliptical copula (Gaussian or t-copula) or an one-parameter Archimedean copula (Gumbel, Frank, Joe or Clayton). Before selecting the copula, an independence test, based on Kendall's tau, is performed on every pair of series using a significance level of . All these estimations have been made using the functions available in the R package *VineCopula* (Schepsmeier (2016); R Core Team (2016)).

Three different estimators are considered and compared to obtain predictions of the future values of the discharge. First of all, we consider the median of the conditional distribution of the discharge given the observed meteorological values (3.10). This can be calculated as the value, , such that,

where the distribution function, , is given in (3.11) if the observed precipitation is zero, or in (3.12) if it is different from zero.

Also, we consider the mean of the conditional distribution (3.10), given the observed values of temperature, humidity, radiation and precipitation in that day. This can be approximated using a Monte Carlo simulation by taking the sample mean of a set of simulated values from (3.10). This is detailed in Appendix B.1.

Finally, we propose a prediction method based on the conditional probability of zero discharge (3.6) and the conditional distribution function of the discharge given the values of the meteorological variables (3.10). Firstly, the conditional probability of discharge is estimated. If this probability is larger than , we consider that the predictive discharge for that day is . If the estimated probability of zero discharge is smaller than , we estimate the mean of the conditional distribution when the discharge is positive in (3.11) and (3.12) using a Monte Carlo simulation as before.

In order to examine the performance of our proposed c-vine model we need two different evaluations, one for the probability of having zero discharge and other for the predicted amount of discharge with the different prediction methods. For the first one we use the Brier Score (Brier (1950)), that measures the distance between the probability and the observation of an event,

— (3.13)

where is the probability that the event will happen and takes value if the event happens and otherwise. For the predictive discharge we will use the Mean Squared Error (MSE) and the Mean Absolute Error (MAE),

— (3.14)

— (3.15)

where is the estimated value and is the true observed value.

3.2 Simulated data

In this Section, we illustrate the proposed methodology with one of the many artificial data that we have generated to examine our procedure. In order to simulate the data we have proposed a model, that is, the number of components and the parameters in each mixture for each marginal distribution, and the structure of each c-vine: the copula family in each node and its parameter. Then, we have estimated the best number of components of each mixture, the parameters for these mixtures, the best copula family for each node in the c-vine structure and their correspondent

Group	Parameter	Temperature Est.	Temperature True	Humidity Est.	Humidity True	Radiation Est.	Radiation True	Precipitation Est.	Precipitation True	Discharge Est.	Discharge True
	p					0.52	0.5				
00	par1.1	0.79	0.8	24.19	23	67.36	7				
	par2.1	0.99	1.0	5.33	5	0.48	0.5				
	par1.2					86.48	86.0				
	par2.2					1	1.0				
	p							0.28	0.6		
01	par1.1	-0.16	-0.2	31.30	31	11.43	12.0	1.43	2.0		
	par2.1	1.04	1.0	1.98	2	0.19	0.2	0.32	2.0		
	par1.2							1.72	1.0		
	par2.2							1.56	0.3		
	p	0.55	0.2								
10	par1.1	0.00	-1.0	15.20	15	13	13.0			2	2
	par2.1	1.32	1.0	4	4	0	0.1			11.54	11
	par1.2	1.17	1.0								
	par2.2	0.84	1.0								
	p										
11	par1.1	1.00	1.0	21.27	24	5.72	6.0	0.99	1.0	1.94	2
	par2.1	0.95	1.0	2.69	3		0.1	0.99	1.0	7.89	8
	par1.2										
	par2.2										

Table 3.1: Comparison between the true and estimated parameters of the mixtures, for the set of simulated data.

parameter, as it is explained in Subsection 3.1.4. Table 3.1 presents the estimation for each parameter in the mixture models. These are compared with the true values of the parameters in order to show the accuracy of the estimations. In the same way, Table 3.2 shows the best family and the value of obtained with the estimations and, also, are compared with the true values. Although, in some cases, the selected family is not the same as the true one, the value of the estimated are very close to the true one. Finally, Fig. 3.3 shows the comparison between the estimated values for all the parameters and the true values used for the simulation of the data.

3.3 Application of the vine copula model

In this Section, our proposed vine copula model is applied to the data provided by GLACKMA from their catchment area in glacier Collins in King George Island. First, the data base is divided in groups according to the seasonality. Second, all the model parameters, of the marginal distributions and the vine copula, are estimated. Third, the conditional probability of having no discharge and the predictive discharge are computed using these parameters of the vine copula

Group	Param.	Family Est.	Family True	Est.	True
00		Frank	Frank	0.38	0.39
		Joe	Joe	0.36	0.36
		Gaussian	Gaussian	-0.21	-0.19
01		Frank	Frank	0.38	0.39
		Gaussian	Gaussian	0.09	0.06
		Joe	Joe	0.13	0.15
		Gaussian	Gaussian	-0.12	-0.13
		Clayton	Clayton	0.21	0.20
		Gaussian	Gaussian	-0.22	-0.19
10		Clayton	Clayton	0.33	0.33
		Frank	Frank	-0.12	-0.11
		Clayton	Clayton	0.21	0.20
		Frank	Frank	-0.25	-0.21
		Clayton	Clayton	0.33	0.33
		Gumbel	Gaussian	0.11	0.13
11		Clayton	Clayton	0.20	0.20
		Frank	Frank	-0.05	-0.09
		Frank	Frank	0.18	0.21
		Gaussian	Gaussian	0.27	0.26
		Gaussian	Gaussian	-0.18	-0.13
		Frank	Frank	0.11	0.11
		Frank	Frank	0.22	0.21
		Gaussian	Frank	-0.11	-0.11
		Gaussian	Gaussian	-0.05	-0.06
		Gaussian	Gaussian	0.12	0.13

Table 3.2: Comparison between the true and the estimated family for the set of simulated data.

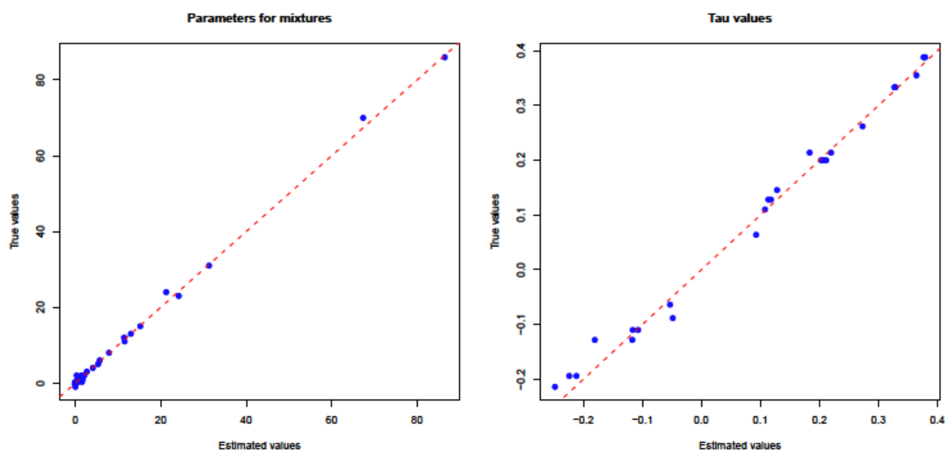

Figure 3.3: Comparison between the parameters estimated by the model for the simulated data and the true ones. The left plot is for the parameters of the mixtures and the right one is for the values in each copula.

Period	Dates	Description
1	26th November - 30th December	Discharge start period. Since the last weeks of spring to early summer. Days can be positive or zero discharge.
2	31st December - 7th April	Main discharge period. Most of the summer. Almost every day has positive discharge.
3	8th April - 15th June	Discharge end period. Since the end of summer and most of autumn. Days can be zero or positive discharge.
4	16th June - 25th November	Zero discharge period. Late autumn, all the austral winter and early spring. There is always zero discharge.

Table 3.3: Distribution of the periods of discharge in King George Island.

model. Finally, the obtained results are compared with those obtained with the bivariate copula model in Chapter 2.

3.3.1 Parameter estimation

Recall that the GLACKMA database consists of five time series of data collected during eleven years. Here, the first ten years are used for parameter identification and data from 10/01/2011 to 12/31/2012 are used for model verification.

First of all, we want to capture the seasonal behaviour of the discharge. In Chapter 2, we try to capture it using partial sums of Fourier terms, but this procedure increases rapidly the number of parameters when we add more meteorological variables. On the other side, Braun (2001) has found three major ablation phases plus a non-ablation phase for each year in glacier behaviour. This suggests us to divide the data in four different periods in order to capture the changes in the relationship between the variables. Table 3.3 shows the different periods selected for this study. As a justification of this division, Fig. 3.4 shows the boxplots of the average daily glacier, grouped by weeks, in the different periods. Apparently, there are different behaviors in the discharge regime. Note that the fourth period has zero discharge in the observed values. Thus, the model will always predict zero discharge in this period, that is, the equation (3.6) will always be zero independently of the values of the other variables because the empirical probability of having no discharge is equal to one.

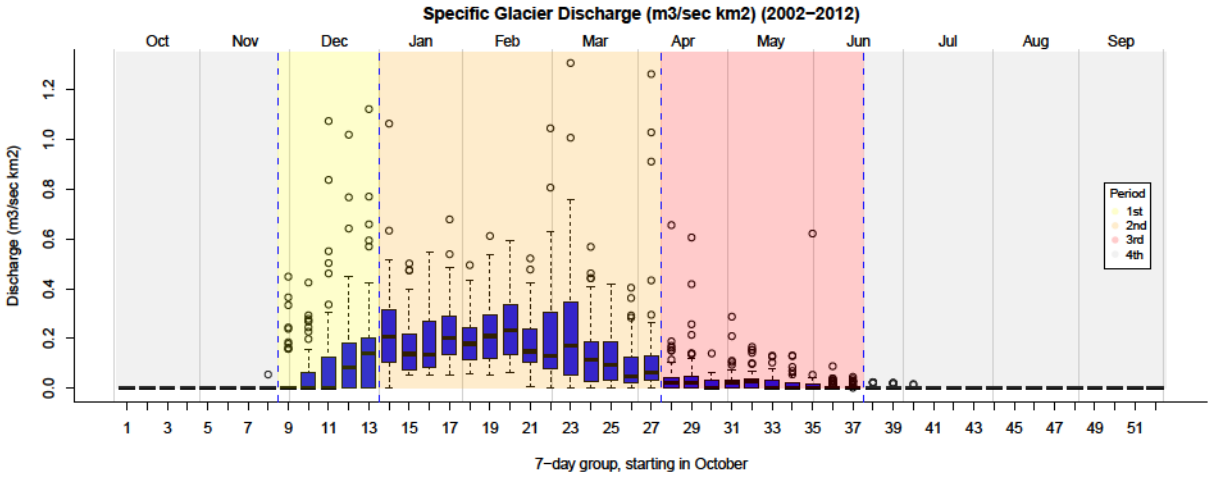

Figure 3.4: Boxplots of the glacier discharge in each week from 2002 to 2012. Different periods are separated by vertical lines and different color shadows.

Firstly, we determine the number of components and the mixture parameters of the marginal functions of the model for the first three periods. As an example, Fig. 3.5 shows the adjustment of the mixtures to the observations of the five variables for the second period and days with positive discharge (the first row for days with positive precipitation and the second one for days with zero precipitation). The number of mixture components is written at the bottom of each plot. An apparently good adjustment between the mixture models and the empirical distributions is observed for all variables and periods. The mixture marginal distribution parameter values are listed in Table 3.4. The first column indicates the period, the second refers to the group , where , respectively, for zero or positive discharge and , respectively, for zero or positive precipitation, and the third shows the number of observations available and used to fit the mixtures.

The following step is to select the copula family and its parameter for each edge in the different vine copula structures. Table 3.6 shows the structure of the c-vine copulas with the value of the parameter for every edge; each row in each edge correspond to one of the first three periods. Note that some edges have the independence copula, denoted by the letter **I**, this means that no significative dependence has been found found in that edge. Different order in the variables

Per	Gr. DP	N	Temperature				
1	00	40		-0.809 (0.191)	1.206 (0.135)		
	01	94		-0.163 (0.111)	1.074 (0.078)		
	10	68	0.220 (0.247)	-0.862 (1.513)	1.133 (0.651)	0.904 (0.170)	0.651 (0.129)
	11	83		0.915 (0.142)	1.295 (0.101)		
2	00	21	0.621 (0.106)	-2.370 (0.234)	0.829 (0.176)	0.665 (0.164)	0.460 (0.115)
	01	16		-1.298 (0.486)	1.944 (0.344)		
	10	283	0.145 (0.056)	-1.954 (0.790)	1.460 (0.430)	1.444 (0.101)	0.937 (0.073)
	11	541	0.177 (0.070)	-0.693 (0.814)	2.189 (0.266)	1.941 (0.082)	1.071 (0.077)
3	00	102		-5.375 (0.378)	3.817 (0.267)		
	01	225	0.665 (0.067)	-6.636 (0.499)	3.523 (0.262)	-0.992 (0.271)	1.230 (0.205)
	10	69		-2.363 (0.288)	2.392 (0.204)		
	11	234	0.517 (0.053)	-3.633 (0.377)	2.759 (0.208)	0.359 (0.107)	0.833 (0.086)
Per	Gr. DP	N	Humidity				
1	00	40		23.527 (5.314)	4.582 (0.99)		
	01	94		31.247 (4.716)	2.421 (0.332)		
	10	68		15.254 (2.648)	3.472 (0.570)		
	11	83		24.086 (3.847)	2.589 (0.379)		
2	00	21		40.476 (12.524)	10.439 (3.173)		
	01	16		16.284 (5.840)	3.326 (1.123)		
	10	283	0.029 (0.019)	6.016 (6.587)	0.240 (0.129)	19.277 (2.065)	3.957 (0.442)
	11	541	0.967 (0.012)	24.565 (1.859)	2.504 (0.192)	305.374 (195.601)	1.829 (0.723)
3	00	102		19.921 (2.829)	3.662 (0.492)		
	01	225		22.324 (2.15)	3.027 (0.271)		
	10	69		14.021 (2.431)	2.783 (0.449)		
	11	234		18.336 (1.761)	2.022 (0.174)		
Per	Gr. DP	N	Radiation				
1	00	40	0.455 (0.09)	69.832 (33.129)	0.605 (0.279)	86.14 (34.047)	1.146 (0.466)
	01	94		12.000 (1.726)	0.184 (0.027)		
	10	68		13.508 (2.287)	0.148 (0.026)		
	11	83		6.229 (0.942)	0.104 (0.016)		
2	00	21		6.980 (2.104)	0.236 (0.074)		
	01	16		2.543 (0.846)	0.103 (0.038)		
	10	283		2.913 (0.232)	0.053 (0.005)		
	11	541		2.721 (0.156)	0.081 (0.005)		
3	00	102	0.591 (0.071)	6.542 (1.492)	2.707 (0.729)	4.308 (1.864)	0.381 (0.142)
	01	225	0.536 (0.150)	4.347 (1.553)	1.945 (0.879)	1.858 (0.456)	0.284 (0.052)
	10	69		1.608 (0.251)	0.184 (0.034)		
	11	234		1.838 (0.157)	0.339 (0.033)		
Per	Gr. DP	N	Precipitation				
1	01	40	0.607 (0.188)	2.06 (0.597)	1.753 (0.896)	1.381 (0.513)	0.277 (0.093)
	11	94		1.242 (0.173)	0.583 (0.099)		
2	01	21		0.731 (0.221)	0.213 (0.090)		
	11	16	0.235 (0.057)	3.487 (0.960)	5.825 (2.337)	1.482 (0.192)	0.326 (0.032)
3	01	102	0.219 (0.051)	7.817 (2.848)	17.953 (7.749)	1.490 (0.186)	0.548 (0.067)
	11	225	0.245 (0.079)	3.503 (1.205)	5.018 (2.611)	1.406 (0.196)	0.354 (0.045)
Per	Gr. DP	N	Discharge				
1	10	40		1.788 (0.283)	11.055 (2.015)		
	11	94		1.640 (0.233)	8.372 (1.390)		
2	10	21		2.247 (0.177)	18.094 (1.593)		
	11	16		1.871 (0.105)	8.525 (0.549)		
3	10	102	0.936 (0.035)	4.029 (0.787)	127.662 (28.612)	18.01 (16.619)	139.819 (121.816)
	11	225	0.355 (0.054)	0.986 (0.143)	9.555 (1.899)	6.744 (1.214)	221.226 (43.461)

Table 3.4: Parameters of the mixtures for all the variables. The first column shows the period of discharge and the second indicates if the group has positive discharge (1 in the first digit) and positive precipitation (1 in the second digit). Third column informs about the number of observed values in each group. The number between parenthesis is the error on the estimation of each parameter.

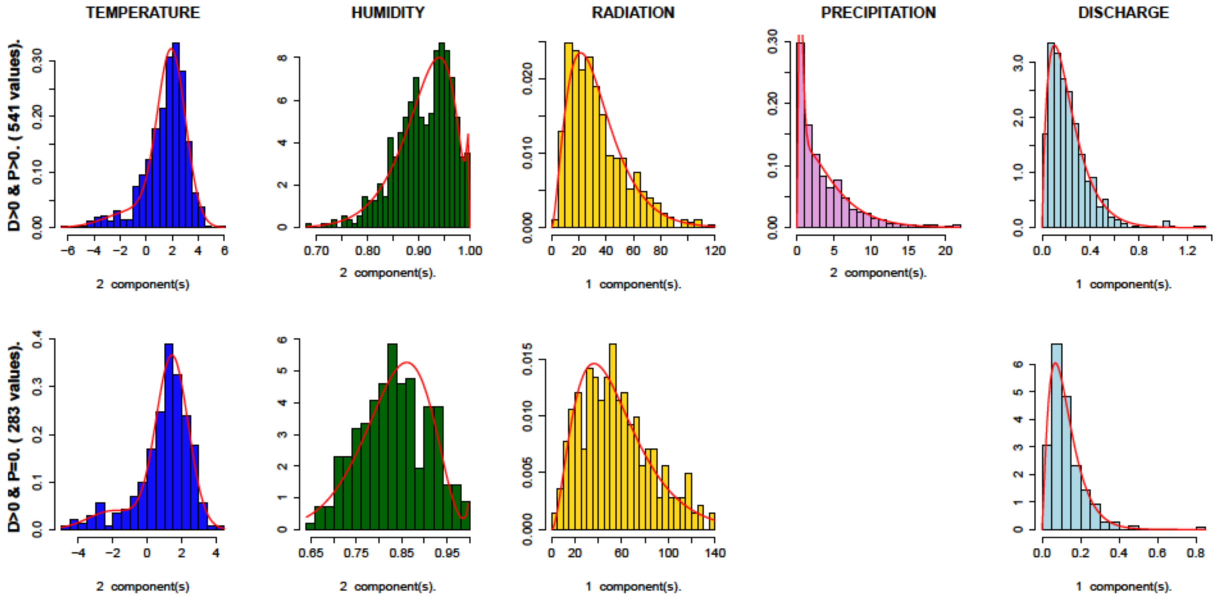

Figure 3.5: Histogram of the observed values, compared with the density function of the adjusted mixture of the correspondent distributions. All histograms are for period 2 and groups of data with positive discharge, with positive precipitation for the first row and without it for the second one. At the bottom is the number of mixture components.

within the c-vine structure has been considered and compared using the BIC criteria and very close values have been found. Also, we have used the Vuong test (Vuong (1989)) to look for differences between different orders but no significative difference has been found. Table 3.5 shows some of the results obtained for the c-vine copula for days with positive discharge and precipitation in the first period. The results for other groups and periods are similar. Then, we have selected the more convenient order to facilitate the evaluation of the probability of discharge and predictive discharge, that is T-H-R-P-D. Also, a goodness-of-fit test has been performed for each of the twelve copulas obtained with the proposed order, which is based on the information matrix equality of White, as detailed in Schepsmeier (2016). Table 3.6 shows the White statistic and the correspondent p-value for each c-vine copula. Observe that both the model and the parameters seem to be appropriate.

In order to examine the goodness of fit of the estimated copulas, we make use of the function (Genest and Rivest (1993)). Fig. 3.7 shows the comparison between the empirical

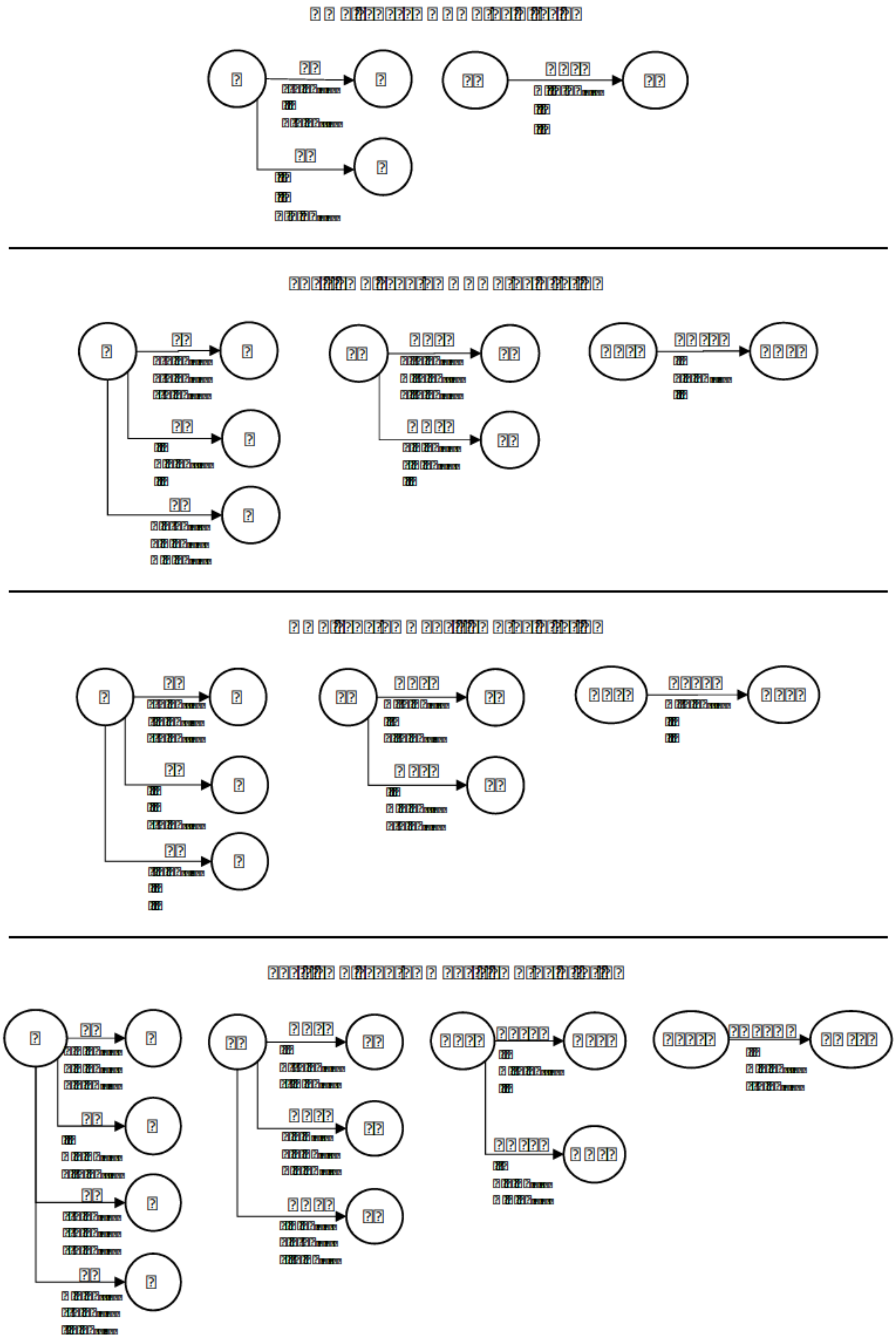

Figure 3.6: Parameters of the c-vine copulas for all periods and groups. Each one of the values in each node is for each period (1,2,3). The copulas are I=Independence, N=Gaussian, C=Clayton, G=Gumbel, F=Frank, J=Joe. The number between parenthesis is the error on the estimation of the parameters.

Order	BIC	Vuong statistic	p-value
THRPD	-19.438	0	1
TDPRH	-19.238	0.190	0.849
HTRPD	-19.194	0.079	0.937
HDRPT	-19.641	-0.042	0.967
RPDTH	-13.809	1.315	0.188
RTDPH	-19.021	0.131	0.896
PTRHD	-13.343	1.191	0.234
PDTRH	-12.591	1.640	0.101
DPRHT	-17.347	0.345	0.730
DTHRP	-20.409	-0.183	0.855

Table 3.5: BIC value of different order combinations for the 5-cvine copula in the first period. Vuong test of comparison with the selected order (THRPD) and the correspondent p-value.

Group	Period 1		Period 2		Period 3	
	White	p-value	White	p-value	White	p-value
00	02.13	0.15	09.24	0.32	08.51	0.60
01	21.53	0.14	19.09	0.60	17.80	0.42
10	19.19	0.79	16.00	0.39	21.76	0.39
11	60.83	0.83	72.90	0.79	73.62	0.06

Table 3.6: White statistic and p-value of the goodness-of-fit test over the twelve c-vine copulas for the selected order.

function for each edge and the theoretical function for the corresponding copula, for the 10 edges of the c-vine copula, for the second period and the group for data with positive values of discharge and precipitation. The dashed lines are bounds corresponding to the independence and comonotonicity copulas, respectively. For the sake of brevity, only one tree has been shown. Apparently, there is good fit between the selected and the empirical copula in all edges for all selected vine structures.

3.3.2 Conditional probability of discharge

Once we have obtained all the model parameters, we are interested in estimating the probability of having zero discharge conditioned to the observed values for the temperature, humidity, radiation and precipitation in each day. As commented above, we will predict positive discharge, with (3.6), for a particular day if the estimated conditional probability of zero discharge is smaller than . Table 3.7 compares predictions with the observed values of the discharge. For the

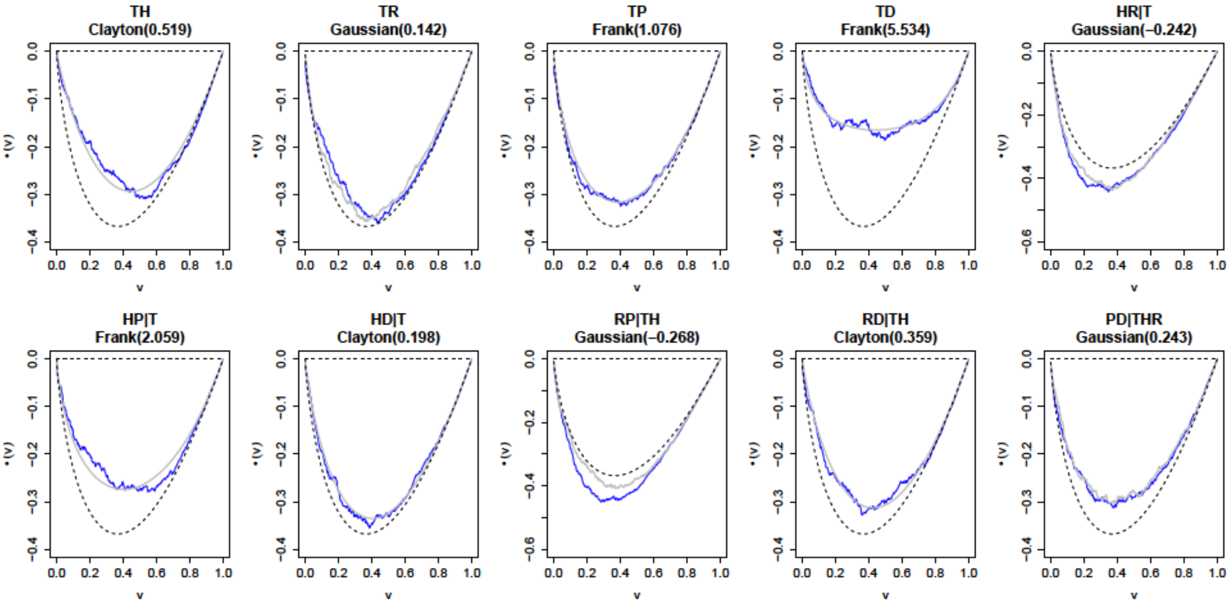

Figure 3.7: Empirical λ-function for the 10 nodes of the period 2 and group where there is discharge and precipitation. The blue line is the empirical and the grey one is the theoretical. The dashed lines in the panels are bounds corresponding to independence and comonotonicity ($\lambda = 0$), respectively.

in-sample data (2002-2011). We obtain the 92.7% of days with zero discharge are correctly predicted with the c-vine model, whereas 88.6% of days with positive discharge are correctly predicted. The performance of the copula model is even better for the out-of-sample data from the last hydrological year, used to validate the model. Our model has a 90.9% and 90.7% of correctly predicted days for days with zero and positive discharge respectively. We have compared these probabilities with the ones obtained with a logistic regression, which has been developed in the same conditions as the vine model, that is, one model for each period. Table 3.8 shows the Brier Score (3.13) for both models, we can see that the vine copula model outperforms the logistic regression, globally and in each period and for the in-sample and the out of sample data. Note that, the smaller the Brier Score the better the predictions.

Additionally, we want to study the dependence of this conditional probability of discharge, on the observed meteorological variables. As an illustration, Fig. 3.8 shows the estimated probability as a function of the temperature for different values of the humidity, in the presence or absence of

	Predicted					
	2002-2011			2011-2012		
			Total			Total
Observed	1883	149	2032	206	21	227
	147	1148	1295	12	127	139

Table 3.7: Comparison between observed discharge and predictions with the vine copula model. On the left, for days with the data used to fit the model (2002-2011). On the right, for the days of the last year (2011-2012), used to validate the model.

	2002-2011		2011-2012	
	Logistic model	Vine model	Logistic model	Vine model
Global	**0.0643**	**0.0607**	**0.0815**	**0.0761**
Period 1	0.1401	0.1314	0.2474	0.2900
Period 2	0.0359	0.0320	0.0254	0.0240
Period 3	0.1999	0.1904	0.1861	0.1463

Table 3.8: Comparison between the Brier Score obtained with a logistic regression and with the vine copula model. On the left, for days with the data used to fit the model (2002-2011). On the right, for the days of the last year (2011-2012), used to validate the model.

precipitation and a fixed value for the radiation. Note that the positive precipitation increases the probability of no discharge, especially when the temperatures are below zero. In these plots we can see, also, that higher temperatures cause a decay in the probability of having no discharge and that an increase of the percentage of humidity increases the probability of having no discharge. Similar plots can be done to compare how the same meteorological conditions in different periods modify the probability of zero discharge.

3.3.3 Predictive discharge

Finally, as described in Subsection 3.1.3, the predictive discharge distribution from (3.10) has been obtained for all days using the three proposed methods explained in Subsection 3.1.4. These predictions have been compared with those obtained in Chapter 2. Table 3.9 shows the MSE (3.14) and the MAE (3.15) obtained for both models. We may observe that in all cases the errors with the vine copula model are smaller than those obtained with the the bivariate copula model. Then, clearly, the vine copula model gives more accurate predictions of the discharge. Finally,

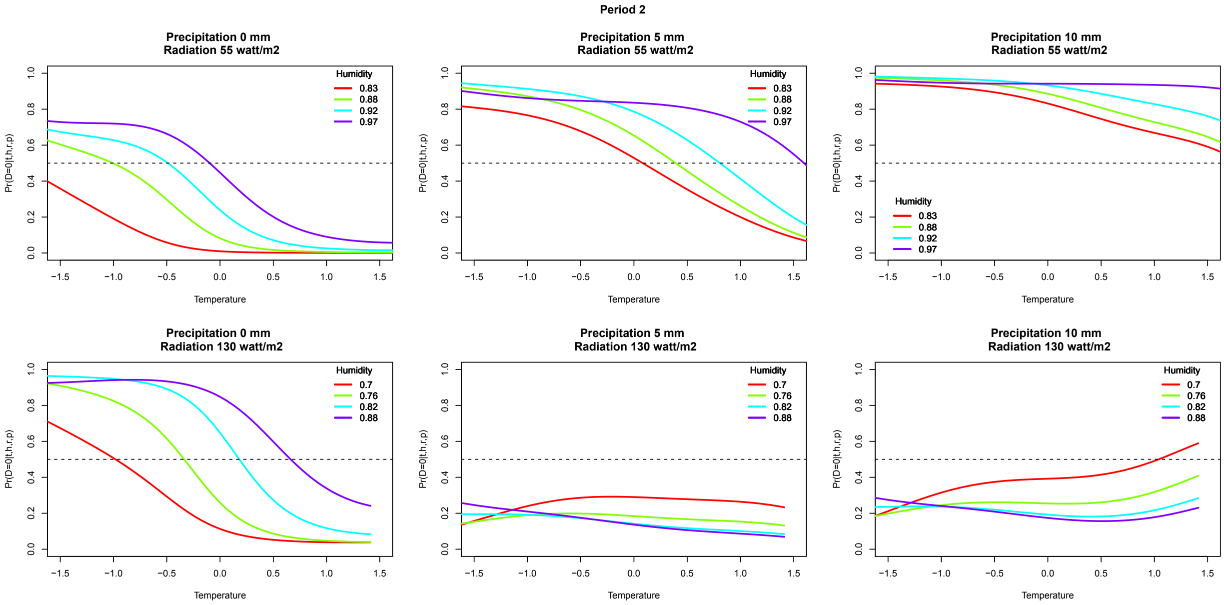

Figure 3.8: Evolution of the probability of having no-discharge during the first period conditioned to different values of the meteorological variables.

Model	Method	2002-2011 SME	2002-2011 MAE	2011-2012 SME	2011-2012 MAE
Vine copula model	Median	0.00621	0.02798	0.01212	0.04682
	Mean	0.00605	0.03175	0.01084	0.04871
	Proposed method	0.00608	0.03031	0.01061	0.04489
Bivariate copula model		0.00718	0.03317	0.03753	0.05362

Table 3.9: Errors of the predicted discharge when vine copula model and bivariate copula model are used. The first two columns have been obtained with the data used to fit the model. The other two have been obtained with the data of the last year, used to validate.

the proposed model is validated with all described methods with the observed values of the discharge of the year (2011-12). The two last columns of Table 3.9 show these measures of the MSE and MAE. It can be observed that the errors of the proposed model are smaller than the ones produced by the bivariate copula model. Therefore, it can be concluded that the use of more meteorological variables in the proposed vine copula model provides more accurate predictions that using simply the temperature as in our previous bivariate copula model.

As an example, the left panel of Fig. 3.9 shows the observed values of the discharge for the year 2005-06 compared with the predictive discharge obtained with the proposed vine copula

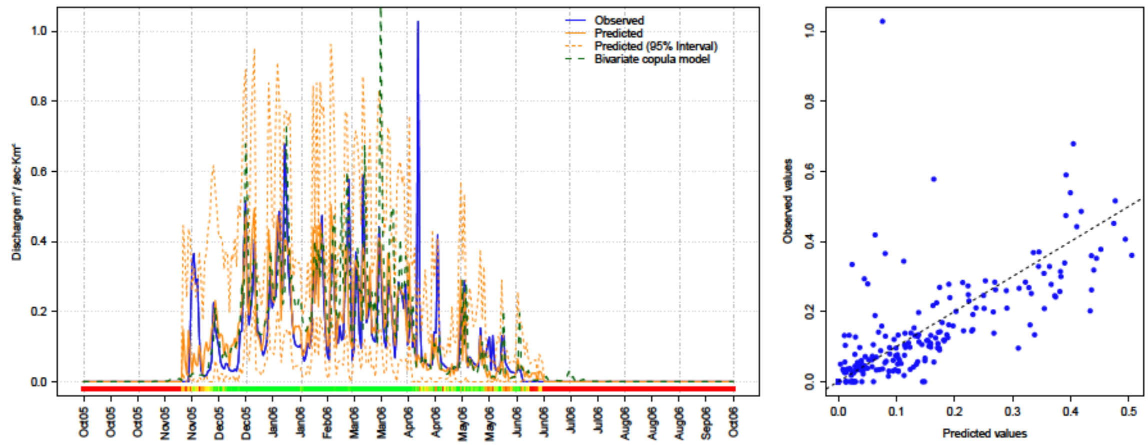

Figure 3.9: Time series of the observed values of the discharge, prediction with c-vine and bivariate copula models and 95% credible intervals for the c-vine model in the year 2005-06. The bottom of the plot shows the conditional probability of discharge of each day in a scale from red (probability zero) to green (probability one). The left panel shows the comparison between the observations and the predictions.

model, together with the corresponding credible intervals, and the predictions obtained with the bivariate copula model. Also, the bottom of the plot illustrates the conditional probability of discharge zero, from red for probability to green for probability . The left panel of Fig. 3.9 shows the scatter plot between the predicted and the observed values. Fig. 3.10 shows similar information for the year 2011-12 whose data were preserved to validate the model. We can see how both plots show a good performance of the proposed vine copula model.

3.4 Conclusion and extensions

In this Chapter, we have proposed a vine copula model for modelling the relationship between the glacier discharge and other meteorological variables, such as, temperature, humidity, solar radiation and precipitation. The probability of zero discharge for each future day is determined given the observed values of the meteorological variables. Also, the predictive value of the discharge is obtained from its conditional distribution given the observations of the meteorological variables. This model has been applied to the data collected by GLACKMA from the glacier Collins between 2002 and 2012. The data base has been divided into four periods and the parameters have been

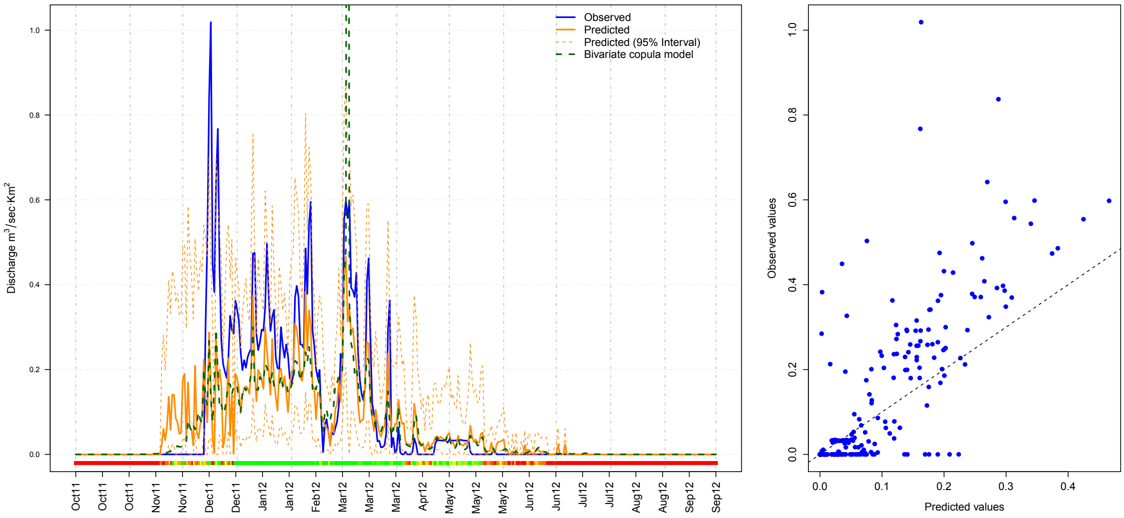

Figure 3.10: Time series of the observed values of the discharge, prediction with c-vine and bivariate copula models and 95% credible intervals for the c-vine model in the year 2011-12 whose data have been used to validate the model. The bottom of the plot shows the conditional probability of discharge of each day in a scale from red (probability zero) to green (probability one). The left panel shows the comparison between the observations and the predictions.

adjusted to obtain the joint distribution of the five variables in each one of these periods. The model shows good performance for all the periods.

Observe that in this work we have assumed a fixed order of the variables. Although different orders produce different c-vine copulas, we have observed that the conditional probability of discharge and the predictive discharge have quite similar results among the different models. However, different vine copula structures and more bivariate copulas could be analyzed in order to achieve better results.

The monitor station in King George island has been registering data that have not been already collected by the GLACKMA association. Our intention is to validate our proposed model with these new data whenever they are available. Moreover, the proposed model could be used in other glaciers whose discharge data is being collected by this association from their Pilot Experimental Watersheds. Furthermore, the model could be used to predict the discharge in glaciers where measuring the real discharge is complex and only the meteorological data is available.

Chapter 4

Hierarchical Vine copula models for the analysis of glacier discharge

The main purpose of this Chapter is to obtain the parameters of the model presented in Chapter 3 under the point of view of the Bayesian Statistic. One possibility could be to follow the same procedure as in Section 2.2, i.e., calculating the likelihood for the model parameters for each period and each group independently, and obtaining a sample of the posterior distribution of the parameters with a Gibbs sampling schema. Instead of this, our thesis now is that the relationship between the same pair of nodes in each c-vine structure is driven by common hyperparameters, regardless of the period or group which they belong to. That is, the dependence between temperature and humidity, for instance, is driven by the same hyperparameters regardless of the season of the year and to the fact of having or not discharge or precipitation. In summary, we propose a hierarchical model over the relationships between the variables of model explained in Chapter 3.

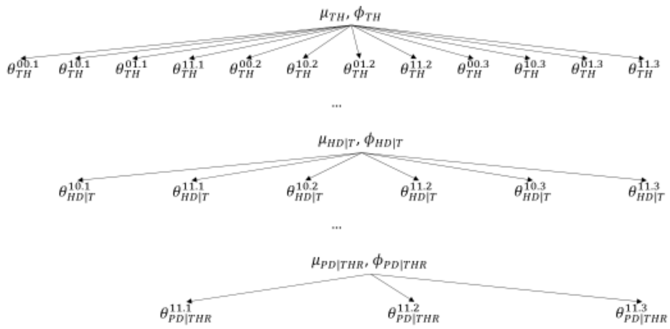

Figure 4.1: Overview of the hierarchical structure of the model, where 3 of the 10 pairs of hyperparameters and their dependent parameters are represented.

4.1 Proposed methodology

In order to reduce the computational cost, we have decided to use a two-step algorithm, that is, firstly we will compute the marginal distribution parameters and then we will obtain the copula parameters. The marginal distribution parameters of each variable has been determined, for example, with classical methods like the ones we have shown in Subsection 3.1.4. So, the values of have already been calculated, where represents all possible edges in the c-vines, and are the combinations of group, which depend on the presence or not of non-zero discharge/precipitation, and period, which are described in Table 3.3[a]. For instance, the superscript corresponds to the group with positive discharge and zero precipitation in the second period.

In the second step, we propose a hierarchical model for each parameter in the c-vine structures of the four groups and the three periods. The hierarchical model together with the Bayesian techniques allow us to face the scarcity of data in some combinations of group and period. Fig. 4.1 shows some examples of the relationships between the parameters and the hyperparameters in our hierarchical model.

Then, we want to sample from the parameter posterior distribution for the c-vine copula of each group and each period. That is, we need to sample all the . As we know the direct

[a]Remember that for the fourth period there is no need of parameters because of the discharge is always zero.

relationship between the copula parameters and the Kendall given the copula family, see Table 1.3, we have decided to sample from the posterior distribution of instead of the posterior distribution of . On the other hand, the values of have support in , if the copula family is Clayton, Gumbel or Joe, and have support in , if the copula family is Gaussian or Frank. Then, we have decided to extend these supports to with a transformation. Then, our model would be,

[b]

where , , and are the functions that transform into , which depend on the copula family selected and are showed in Table 1.3.

4.1.1 RWMH algorithm

Firstly, we develop a RWMH algorithm to sample from the posterior distribution of the . We have selected a 2-step algorithm:

1. Sample
2. Sample

For the first step we have a Normal-Gamma distribution, then we know the conjugate

[b] ( ) if the ∈ like in Clayton, Gumbel and Joe copulas.

posterior distribution and we can sample directly from it:

— — where

with is the mean of , is the standard deviation of and or depending on the number of parameters associated to each pair of hyperparameters.

For the second step, we propose a Gibbs sampling schema which is carried out by cycling repeatedly through draws of each parameter conditional on the remaining parameters (Tierney (1994)). In particular, we select a simple one-dimensional RWMH where each model parameter is updated separately using normal candidate distributions whose mean is given by the previous value of each parameter in the algorithm and whose variance can be calibrated to obtain good acceptance rates. The parameter prior distributions for the model are needed, since the parameters depend on the values of the hyperparameters obtained in the previous step we need the induced priors, which must be obtained with the transformation function, , as

in our case we have two transformation functions, which depend on the range of values of the :

Then the induced priors will be,

Finally, the likelihood of the c-vine copula is computed as the sum of the likelihood of all the bivariate copulas in its structure. Once we have obtain a parameter for all the copulas in the different c-vine structures, we transform them into the , with the logit transformation, and use them to actualize the value of the hyperparameters on the first step. The steps 1 and 2 are executed repeatedly until the convergence is achieved and the MCMC has the desired length.

4.1.2 ABC algorithm

Now, we develop an Approximate Bayesian Computation algorithm (ABC) to sample from the posterior distribution of the . We consider for our work the algorithm proposed by Beaumont et al. (2002), which has been implemented in the R-package `abc` (Csilléry et al. (2012)). So, our algorithm will be as follows: Firstly, we need the summary statistics of the observations to compare them with the ones obtained with the proposed parameters. Although they are not sufficient statistics, we have chosen the Kendall rank correlation between each pair of nodes in the c-vine structures. Then, we have a vector with these correlations. For the proposed hyperparameters, we simulate a big amount of 10-dimensional vectors , and the same amount of 10-dimensional vectors , where

, .

For each pair , we simulate 12 values , these are transformed into with the logit transformation, and the are transformed into with the corresponding function from Table 1.3, and the apropriate copula family. Observe that we do not use the 12 values in every case considering that not all the c-vines have the same number of nodes. Table 4.1 shows in which edges of the c-vines the parameters are needed.

Now we have the c-vine structures with the proposed parameters , that corresponds with the columns of the Table 4.1. In each one of these c-vines we can simulate values and, then, we can obtain the Kendall rank correlation between each pair of c-vine nodes in each simulation, that is, the summary statistics. Then, we have ( ) , the proposed parameters in step i ( ) , and the summary statistics of the simulations obtained with these parameters . The Euclidean distance between and each is calculated: . The subset of whose Euclidean distance to is smaller than a tolerance threshold, , is selected. We denote as the index of the selected parameters. Following the idea of Beaumont et al. (2002), we perform a local linear regression method in order to correct the imperfect match between and . Then, we need to minimize the expression,

where is a vector of linear regression coefficients. Simulations whose distance to is smaller are given more weight, , where is the Epanechnikov kernel. Finally, a sample from the posterior parameter distribution is obtained by correcting the with the expression:

finally, will be sample of the posterior distribution of the hyperparameters.

4.2 Simulated data

In this Section, we compare the proposed RWMH and ABC algorithms with an artificial sample of a hierarchical model of four groups, of 3, 4, 4 and 5 c-vines respectively and three periods. Table 4.2 shows some of the results obtained with both methods. In particular the ones obtained for one

		Period 1				Period 2				Period 3			
		00	01	10	11	00	01	10	11	00	01	10	11
		x	x	x	x	x	x	x	x	x	x	x	x
		x	x	x	x	x	x	x	x	x	x	x	x
			x		x		x		x		x		x
				x	x			x	x			x	x
		x	x	x	x	x	x	x	x	x	x	x	x
			x		x		x		x		x		x
				x	x			x	x			x	x
			x		x		x		x		x		x
				x	x			x	x			x	x
					x				x				x

Table 4.1: Relation between hyperparameters and needed —values in each edge, period and group.

		RWMH	ABC
	True value	mean (cred int.)	mean (cred int.)
	0.317	0.324 (0.297,0.353)	0.279 (0.232,0.322)
	0.561	0.565 (0.524,0.606)	0.528 (0.504,0.550)
	0.776	0.783 (0.771,0.795)	0.751 (0.706,0.794)
	-0.193	-0.235 (-0.299,-0.172)	-0.209 (-0.247,-0.173)
	0.675	0.670 (0.643,0.695)	0.632 (0.596,0.667)
	0.622	0.621 (0.579,0.661)	0.638 (0.600,0.677)

Table 4.2: Comparison between the true values and the ones obtained both algorithms, for one of the 4-node c-vines in the second period. The first value is the true value, the second one is the mean of the MCMC and finally there is the credible interval. Results for simulated data.

of the 4-node c-vine in the second period. Fig. 4.2 shows the comparison of the posterior samples of the parameters obtained with both methods. Apparently, the results on the ABC method are closer to the true values than the ones obtained with the RWMH algorithm. Also we have measure the execution time of both methods. Table 4.3 shows some of the execution times obtained for both algorithms[c] over the simulated data. The ABC algorithm is, approximately, a faster than RWMH algorithm. This test has been done with a desktop computer with a Intel(R) Core(TM) i5-330M CPU@2.60GHz processor.

We have used these simulated data with the algorithm of the previous Chapter with the aim

[c]The RWMH algorithm needs 2,000 iterations to achieve convergence.

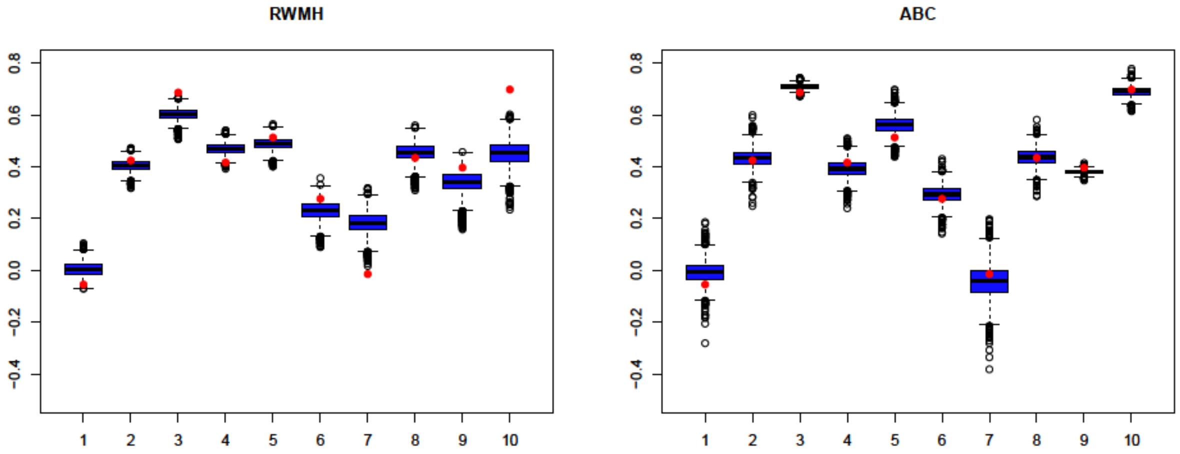

Figure 4.2: Boxplot of the sample of the parameter posterior distribution for the 5-node c-vine of the second period in the hierarchical model. Results for simulated data.

		Time	
Length	Acep./tol.	RWMH	ABC
10,000			
30,000			
50,000			

Table 4.3: Execution times of RWMH and ABC algorithms, obtained with a desktop computer with a Intel(R) Core(TM) i5-330M CPU@2.60GHz processor, for different lengths of the samples of the parameter posterior distributions. Results for simulated data.

of comparing, as far as possible, both methods. Since in the case of classical inference we have a point estimation, we have compared the values with the means of the chains obtained, with the algorithm ABC and with the true values of the tau parameters. In almost all cases the average of the chains is closer to the true value of the tau for the ABC-hierarchical algorithm. In addition, the mean value of the distance from the estimated parameters to the true ones is 0.110 for the point estimation algorithm and 0.059 for the ABC-hierarchical algorithm.

Also, these simulated data have been used with an ABC algorithm, but assuming that the model in not hierarchical. The distances to the true values are larger than the ones of the hierarchical model, specially for groups with few data. The mean value of the distance from

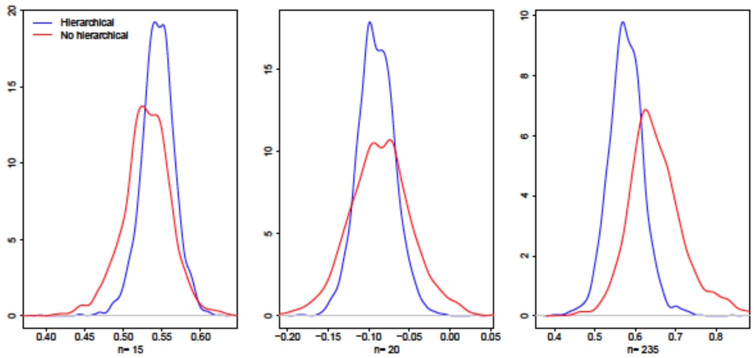

Figure 4.3: Density of the posterior sample obtained with a hierarchical and a non-hierarchical ABC algorithms. The number below each plot is the number of observations used in the algorithms. Results for simulated data from a hierarchical model.

the estimated parameters to the true ones is 0.066 for the ABC-non-hierarchical algorithm and 0.059 for the ABC-hierarchical algorithm. As an example, Fig. 4.3 shows the density plot of the posterior distribution sample of 3 of the 75 copula parameters. We can see that the uncertainty is smaller when we assume the hierarchical model.

4.3 Application of the ABC algorithm

In this Section, we apply the proposed ABC methodology to the data provided by GLACKMA from their catchment area in glacier Collins in King George Island. As it is explained in Section 4.1, the parameter for the marginal functions have been previously calculated as in Section 3.3. Also, we suppose that the copula families for each edge in the c-vine structures are known. Then, we only have to obtain the sample of the posterior parameter distribution for the values of each edge.

Table 4.4 shows the mean, posterior deviation and credible intervals for the copula parameters in each edge of the c-vines that results after applying the ABC algorithm over the glacier observed data. Note that the empty cells correspond to edges where the independence copula has been placed.

As in the previous Chapter, we have obtained the probability of having zero discharge given

	F	3.95(2.13)	(1.01,8.08)	J	1.94(0.31)	(1.44,2.63)	N	0.42(0.04)	(0.34,0.49)
	J	1.27(0.06)	(1.15,1.38)	N	0.11(0.12)	(-0.11,0.34)	N	0.31(0.02)	(0.26,0.35)
	N	-0.39(0.02)	(-0.42,-0.35)	F	-2.84(0.57)	(-4.04,-1.80)	F	-0.83(0.10)	(-1.02,-0.64)
	F	3.82(1.23)	(1.73,6.58)	J	2.23(0.12)	(2.01,2.47)	F	3.30(0.15)	(3.01,3.61)
	N	-0.13(0.02)	(-0.16,-0.09)	F	-1.15(0.11)	(-1.37,-0.94)	F	1.46(0.56)	(0.42,2.57)
	J	1.63(0.13)	(1.39,1.91)	C	0.48(0.06)	(0.35,0.60)	J	1.23(0.27)	(1.00,1.90)
	N	-0.19(0.15)	(-0.48,0.12)	N	-0.19(0.02)	(-0.23,-0.15)	F	-1.23(0.10)	(-1.43,-1.04)
	C	0.51(0.16)	(0.21,0.84)	N	0.74(0.01)	(0.73,0.75)	C	0.93(0.44)	(0.24,1.88)
	N	-0.35(0.02)	(-0.39,-0.29)	N	-0.15(0.05)	(-0.24,-0.05)	F	-0.33(0.28)	(-0.89,0.21)
	C	1.45(0.07)	(1.30,1.59)	C	0.71(0.08)	(0.56,0.86)	F	1.81(0.37)	(1.11,2.57)
	F	-0.94(0.04)	(-1.02,-0.85)	C	0.46(0.15)	(0.18,0.80)	J	1.39(0.07)	(1.27,1.53)
	C	0.64(0.29)	(0.14,1.30)	F	4.57(0.86)	(3.03,6.54)	G	1.40(0.09)	(1.24,1.60)
	F	-2.22(0.24)	(-2.71,-1.73)	N	-0.35(0.01)	(-0.38,-0.32)	F	-3.73(0.40)	(-4.60,-2.97)
	C	0.78(0.19)	(0.43,1.16)	F	1.18(0.22)	(0.73,1.63)	N	-0.08(0.20)	(-0.46,0.31)
	N	0.19(0.08)	(0.03,0.34)	F	2.10(0.45)	(1.19,3.03)	F	0.36(0.37)	(-0.39,1.08)
	C	0.91(0.14)	(0.65,1.21)	C	0.59(0.25)	(0.14,1.11)	F	2.75(0.07)	(2.61,2.90)
	F	0.07(0.29)	(-0.49,0.60)	N	0.02(0.06)	(-0.09,0.13)	F	-0.89(0.51)	(-1.96,0.09)
	F	2.71(0.28)	(2.21,3.26)	F	1.02(0.49)	(0.06,1.99)	F	1.63(0.11)	(1.43,1.85)
	N	0.37(0.01)	(0.35,0.39)	F	5.23(0.49)	(4.31,6.29)	J	1.94(0.05)	(1.84,2.04)
	N	-0.18(0.04)	(-0.26,-0.10)	N	-0.20(0.02)	(-0.25,-0.15)	F	-1.85(0.11)	(-2.06,-1.64)
	F	1.49(0.32)	(0.86,2.14)	F	1.79(0.13)	(1.53,2.06)	C	0.50(0.06)	(0.39,0.62)
	F	1.69(0.15)	(1.38,1.98)	C	0.41(0.20)	(0.05,0.86)	F	-0.56(0.12)	(-0.81,-0.31)
	F	-1.13(0.22)	(-1.58,-0.71)	N	-0.30(0.07)	(-0.43,-0.16)	N	-0.02(0.07)	(-0.15,0.12)
	N	-0.07(0.05)	(-0.17,0.02)	C	0.59(0.17)	(0.27,0.96)	N	0.23(0.02)	(0.18,0.28)
	N	0.11(0.11)	(-0.13,0.33)	N	0.24(0.03)	(0.18,0.29)	F	0.82(0.31)	(0.23,1.43)

Table 4.4: Model parameters for the copulas in each c-vine structure. Each parameter is obtained as the mean of the parameter posterior sample. The posterior deviation is the number between parenthesis. The third column of each parameter is the credible interval. These results are for the glacier observed data.

the observed values of the other meteorological variables. In this case, one probability is obtained for every value in the parameters posterior sample and every day, and then, the mean or the median of these probabilities for each day has been obtained. Table 4.5 shows the comparison between the observed and predicted discharge. As one can see, the results are very similar to the obtained with the classical inference. Now, we could plot the evolution of the probability of having zero discharge conditioned to different values of the meteorological variables, but we can add to these evolutions the credible intervals provided by the Bayesian point of view. Fig. 4.4 shows the evolution of these probabilities.

Furthermore, values of the predictive discharge conditioned to the other meteorological variables are calculated with the posterior parameter sample. Then, we can obtain the mean or the median for each day, but also, similarly as we do in Section 3.3.3, we propose a prediction

	Predicted					
	2002-2011			2011-2012		
			Total			Total
Observed	1894	138	2032	207	20	227
	225	1070	1295	32	107	139

Table 4.5: Comparison between observed discharge and predictions with the vine copula model. On the left, for days with the data used to fit the model (2002-2011). On the right, for the days of the last year (2011-2012), used to validate the model.

	2002-2011			2011-2012		
	Bayesian	Classical		Bayesian	Classical	
	Vine model	Logistic model	Vine model	Vine model	Logistic model	Vine model
Global	**0.0771**	**0.0643**	**0.0607**	**0.1106**	**0.0815**	**0.0761**
Period 1	0.1431	0.1401	0.1314	0.3041	0.2474	0.2900
Period 2	0.0514	0.0359	0.0320	0.0720	0.0254	0.0240
Period 3	0.2451	0.1999	0.1904	0.2271	0.1861	0.1463

Table 4.6: Comparison between the Brier Score for the conditional probability obtained with Bayesian inference compared with the one obtained with classical inference (logistic regression and vine copula model). On the left, for days with the data used to fit the model (2002-2011). On the right, for the days of the last year (2011-2012), used to validate the model.

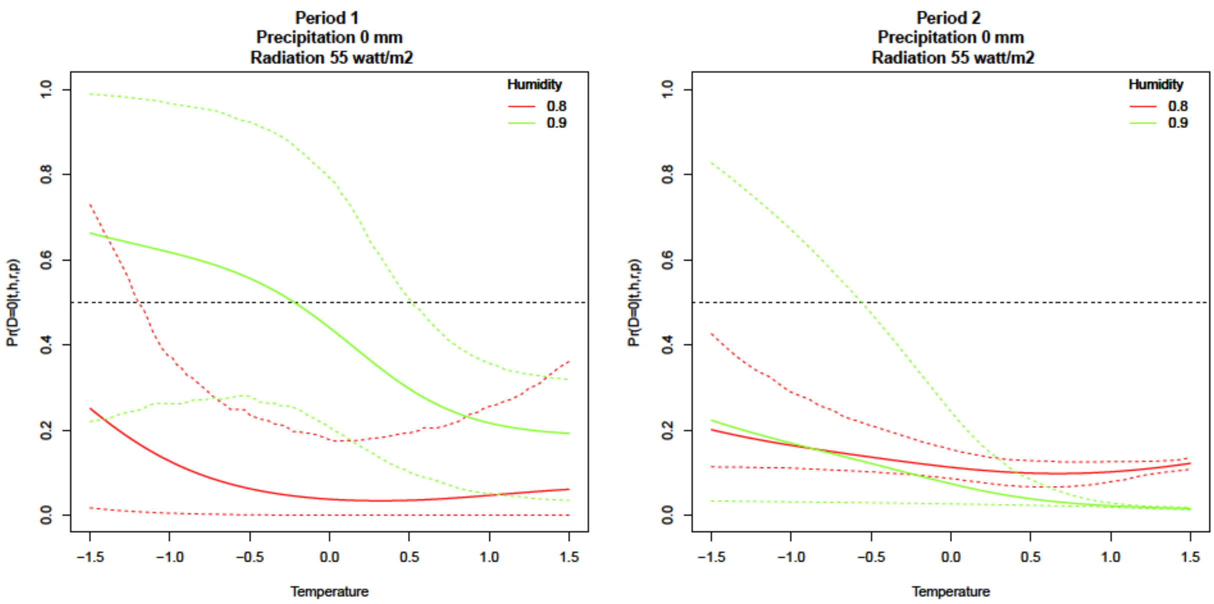

Figure 4.4: Evolution of the probability of having no-discharge during the first period conditioned to different values of the meteorological variables. Dashed lines corresponds to the credible intervals of these probabilities.

Model	Method	2002-2011		2011-2012	
		SME	MAE	SME	MAE
	Mean	0.00813	0.04344	0.01016	0.04949
Vine copula model	Median	0.00772	0.03985	0.01059	0.04892
	Prop. method	0.00756	0.03646	0.01110	0.04238

Table 4.7: Errors of the predicted discharge when c-vine model, estimated with the ABC method, is used. The first two columns have been obtained with the data used to fit the model. The other two have been obtained with the data of the last year, used to validate the model.

method that assigns a zero value if the probability of zero discharge is greater than or the correspondent value if this probability is less. This is done with every element of the samples. Finally, we can obtain the mean or the median for every day. Table 4.6 shows the Brier Score (3.13) for the conditional probability obtained with the ABC algorithm, the results are compared with the ones obtained in the previous Chapter. Table 4.7 shows the MSE (3.14) and the MAE (3.15) when the predictions of the model are compared with the observed values.

Fig. 4.5 shows the predictive discharge compared with the observed values for the year 2005-06. The predicted values for the discharge, in the plot at the left, have been obtained with the mean whereas the predictions on the right plot have been estimated with the proposed method. At the bottom of each plot we have plot the conditional probability of zero discharge for each day in a scale from red (probability zero) to green (probability one). The proposed method seems to be more conservative and provides predictive discharge smaller the one obtained with the mean in periods 1 and 3. For period 2 the predictions are almost the same in both methods because the probability of discharge zero is smaller than .

4.4 Conclusion and extensions

In the previous Chapter, we proposed a model that divided the data into periods of time through the year and into groups, depending on the presence or not of discharge and/or precipitation, and used c-vine copulas to model the relationship within each combination of period-group independently of each other. In this Chapter, we use the same structure to search a model for the

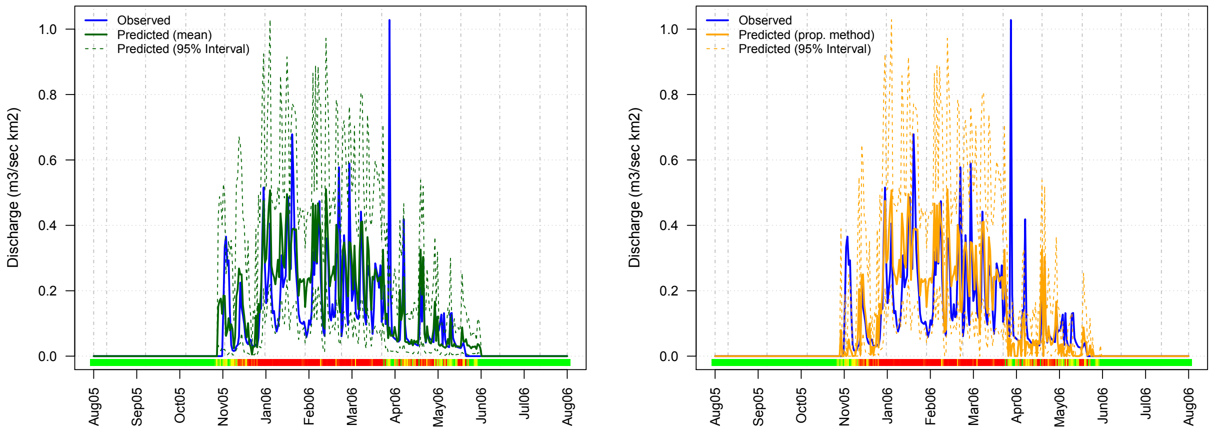

Figure 4.5: Time series of the observed values of the discharge, prediction with c-vine model and 95% credible intervals in the year 2005-06. On the left for the mean and on the right for the proposed method. The bottom of each plot shows the conditional probability of zero discharge for each day in a scale from red (probability zero) to green (probability one).

data, but now, we have supposed that the relationships between the different nodes in the c-vine structures are not independent through the time or the groups, but they come from distributions with common hyperparameters. Therefore, we have proposed a hierarchical model in which the relationships between the same nodes in the different c-vines come from the same distribution. For instance, the relationship between temperature and humidity in a group where there is neither discharge nor precipitation is related to the relationship between temperature and humidity in a group with positive discharge and precipitation.

Furthermore, we wanted to make inference over the parameters of this model using the Bayesian point of view. We have compared two algorithms to obtain samples of the parameter posterior distributions. Every iteration of the first one has two recursive steps. In the first step, the hyperparameters are updated directly from the conjugate of a Normal-Gamma distribution, with the last values of the parameters. In the second step, the parameters are updated with a RWMH algorithm in which the parameter prior distributions change, depending on the last values of the hyperparameters. The second algorithm is based on Approximate Bayesian Computation

(ABC), improved with a weighted local linear regression. In this algorithm, random sets of hyperparameters and parameters are compared, via summary statistics, with the observed data, in order to select the subset of the most similar parameters to the real ones. The results obtained with both algorithms are quite similar, but the ABC algorithm is faster.

The ABC algorithm has been applied to the GLACKMA database. The results are quite similar those ones obtained in Chapter 3, but now we have a sample from the parameter posterior distributions and, for instance, we can add the credible intervals both to the probability of discharge and to the predictive discharge for each of the days.

Other variations of the ABC algorithms appear in the literature and they could be adapted to our hierarchical model and compared the execution times and the obtained results. Another possibility would be to integrate the parameter of the mixture distribution functions of the variables into the hierarchical model, but in this case the amount of variables will increase considerably. Also, more vine structures, like the so called Regular-vine copulas, which include the c-vine copulas, could be considered. Also, It will be very interesting to include more variables such as the direction and speed of the wind. There are studies that show how they have influence on glacier melting in the Antarctica (Orr et al. (2008)).

Comparison between the different glaciers monitored by GLACKMA could be done analysing the correspondent model parameters of each one.

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

Acar, E. F., Genest, C., and NešLehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90.

Ausin, M. C. and Lopes, H. F. (2010). Time-varying joint distribution through copulas. *Computational Statistics & Data Analysis*, 54(11):2383–2399.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

Baranowski, S. and Jurasz, K. (1977). *The subpolar glaciers of Spitsbergen seen against the climate of this region*. Wydawnictwa Uniwersytetu Wrocławskiego.

Barrand, N., Vaughan, D., Steiner, N., Tedesco, M., Kuipers Munneke, P., Broeke, M., and Hosking, J. (2013). Trends in Antarctic Peninsula surface melting conditions from observations and regional climate modeling. *Journal of Geophysical Research: Earth Surface*, 118(1):315–330.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual review of ecology, evolution, and systematics*, 41:379–406.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268.

Bers, A. V., Momo, F., Schloss, I. R., and Abele, D. (2013). Analysis of trends and sudden changes in long-term environmental data from King George Island (Antarctica): relationships between global climatic oscillations and local system response. *Climatic change*, 116(3-4):789–803.

Braun, M. (2001). Ablation on the ice cap of King George Island (Antarctica). *University of Freiburg (Germany)*.

Braun, M., Simões, J. C., Vogt, S., Bremer, U. F., Saurer, H., and Aquino, F. E. (2002). Satellite image map of King George Island, Antarctica. Supplement to: Braun, M et al. (2002): A new satellite image map of King George Island (South Shetland Islands, Antarctica). Polarforschung, 71(1-2), 47-48, hdl:10013/epic.29872.d001.

Brechmann, E., Czado, C., and Paterlini, S. (2014). Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *Journal of Banking & Finance*, 40:271–285.

Brechmann, E. C. and Schepsmeier, U. (2013). Modeling dependence with C- and D-Vine Copulas: The R package CDVine. *Journal of Statistical Software*, 52(3):1–27.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3.

Burda, M. and Prokhorov, A. (2014). Copula based factorization in Bayesian multivariate infinite mixture models. *Journal of Multivariate Analysis*, 127:200–213.

Cogley, J., Hock, R., Rasmussen, L., Arendt, A., Bauder, A., Braithwaite, R., Jansson, P., Kaser, G., Möller, M., Nicholson, L., et al. (2011). Glossary of glacier mass balance and related terms. *IHP-VII technical documents in hydrology*, 86:965.

Cong, R. G. and Brady, M. (2012). The interdependence between rainfall and temperature: copula analyses. *The Scientific World Journal*, 2012.

Csilléry, K., François, O., and Blum, M. G. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution*, 3(3):475–479.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. académie royale de Belgique. *Bulletin de la Classe des Sciences*, 65(5):274–292.

Deheuvels, P. (1981). A Kolmogorov-Smirnov type test for independence and multivariate samples. *Revue roumaine de mathématiques pures et appliquées*, 26(2):213–226.

Domínguez, M. (2004). Software for gauging. In *VI Symposium Glacier Caves and Karst in Polar Regions. Ny-Alesund (Svalbard), Norway*, pages 27–36.

Domínguez, M. and Eraso, A. (2007). Substantial changes happened during the last years in the icecap of King George, Insular Antarctica. *Karst and Criokarst. Studies of the*, 45:87110.

Domínguez, M., Eraso, A., and Lluberas, A. (2004). Annual wave of glacier discharge in the collins subpolar icecap in King George island. In *VI Symposium Glacier Caves and Karst in Polar Regions, Ny-Alesund (Svalbard), Norway*, pages 89–108.

Dos Santos Silva, R. and Lopes, H. F. (2008). Copula, marginal distributions and model selection: a Bayesian note. *Statistics and Computing*, 18(3):313–320.

Durante, F. and Sempi, C. (2010). Copula theory: an introduction. In *Copula theory and its applications*, pages 3–31. Springer.

Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650.

Embrechts, P., Frey, R., and McNeil, A. (2005). Quantitative risk management. *Princeton Series in Finance, Princeton*, 10.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.

Eraso, A. and Pulina, M. (1994). *Cuevas en hielo y ríos bajo los glaciares*. GB2430. A9. E72 1992. McGraw-Hill. Madrid.

Erhardt, V. and Czado, C. (2012). Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuarial Journal*, 2012(2):106–129.

Gaenssler, P. et al. (2013). *Seminar on empirical processes*, volume 9. Birkhäuser.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043.

Gómez, M., Ausín, M. C., and Domínguez, M. C. (2016). Vine copula models for predicting water flow discharge at King George Island, Antarctic. *UC3M Working Papers*, Statistics and Econometrics 16–12.

Gómez, M., Ausín, M. C., and Domínguez, M. C. (2017). Seasonal copula models for the analysis of glacier discharge at King George Island, Antarctica. *Stochastic Environmental Research and Risk Assessment*, 31(5):1107–1121.

Gyasi-Agyei, Y. (2013). Evaluation of the effects of temperature changes on fine timescale rainfall. *Water Resources Research*, 49(7):4379–4398.

Gyasi-Agyei, Y. and Melching, C. S. (2012). Modelling the dependence and internal structure of storm events for continuous rainfall simulation. *Journal of Hydrology*, 464:249–261.

Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction-simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Hamlet, A. F. and Lettenmaier, D. P. (1999). Effects of climate change on hydrology and water resources in the Columbia River Basin.

Hock, R. (1999). A distributed temperature-index ice-and snowmelt model including potential direct solar radiation. *Journal of Glaciology*, 45(149):101–111.

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, 282(1):104–115.

Hock, R. (2005). Glacier melt: a review of processes and their modelling. *Progress in physical geography*, 29(3):362–391.

Jansson, P., Hock, R., and Schneider, T. (2003). The concept of glacier storage: a review. *Journal of Hydrology*, 282(1):116–129.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Killiches, M., Kraus, D., and Czado, C. (2017). Examination and visualisation of the simplifying assumption for vine copulas in three dimensions. *Australian & New Zealand Journal of Statistics*, 59(1):95–117.

Kole, E., Koedijk, K., and Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8):2405–2423.

La Frenierre, J. and Mark, B. G. (2014). A review of methods for estimating the contribution of glacial meltwater to total watershed discharge. *Progress in Physical Geography*, 38(2):173–200.

Manner, H. and Reznikova, O. (2012). A survey on time-varying copulas: specification, simulations, and application. *Econometric Reviews*, 31(6):654–687.

Marsh, P. (1999). Snowcover formation and melt: recent advances and future prospects. *Hydrological Processes*, 13(14-15):2117–2134.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Ohmura, A. (2001). Physical basis for the temperature-based melt-index method. *Journal of Applied Meteorology*, 40(4):753–761.

Orr, A., Marshall, G. J., Hunt, J. C., Sommeria, J., Wang, C.-G., Van Lipzig, N. P., Cresswell, D., and King, J. C. (2008). Characteristics of summer airflow over the antarctic peninsula in response to recent strengthening of westerly circumpolar winds. *Journal of the Atmospheric Sciences*, 65(4):1396–1413.

Patton, A. J. (2009). Copula–based models for financial time series. In *Handbook of financial time series*, pages 767–785. Springer.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., and Corripio, J. (2005). An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland. *Journal of Glaciology*, 51(175):573–587.

Pitt, M., Chan, D., and Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, pages 537–554.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Romera, R. and Molanes, E. M. (2008). Copulas in finance and insurance. *Statistics and Econometrics Series 21*.

Rosenberg, J. V. and Schuermann, T. (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial economics*, 79(3):569–614.

Rückamp, M., Braun, M., Suckro, S., and Blindow, N. (2011). Observed glacial changes on the King George island ice cap, Antarctica, in the last decade. *Global and Planetary Change*, 79(1):99–109.

Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric theory*, 20(03):535–562.

Scaillet, O. and Fermanian, J.-D. (2002). Nonparametric estimation of copulas for time series. *Journal of risk*, 5(03):55–54.

Schepsmeier, U. (2010). *Maximum likelihood estimation of C-vine pair-copula constructions based on bivariate copulas from different families*. PhD thesis, Masters thesis, Technische Universität München.

Schepsmeier, U. (2016). A goodness-of-fit test for regular vine copula models. *Econometric Reviews*, pages 1–22.

Schoelzel, C. and Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. *Nonlin. Processes Geophys.*, 15(5):761–772.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons.

Sicart, J. E., Hock, R., and Six, D. (2008). Glacier melt, air temperature, and energy balance in different climates: The Bolivian Tropics, the french Alps, and northern Sweden. *Journal of Geophysical Research: Atmospheres*, 113(D24).

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, (8):229–231.

Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479.

Song, P. X.-K., Fan, Y., and Kalbfleisch, J. D. (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, 100(472):1145–1158.

Spanhel, F. and Kurz, M. S. (2015). Simplified vine copula models: Approximations based on the simplifying assumption. *arXiv preprint arXiv:1510.06971*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145(2):505–518.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.

Toews, M. W., Whitfield, P. H., and Allen, D. M. (2007). Seasonal statistics: The seas package for R. *Computers & geosciences*, 33(7):944–951.

Turner, B. M. and Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, 56(2):69–85.

Turner, B. M. and Van Zandt, T. (2014). Hierarchical approximate Bayesian computation. *Psychometrika*, 79(2):185–209.

Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A., and Iagovkina, S. (2005). Antarctic climate change during the last 50 years. *International journal of Climatology*, 25(3):279–294.

Vaughan, D., Marshall, G., Connolley, W., Parkinson, C., Mulvaney, R., Hodgson, D., King, J., Pudsey, C., and Turner, J. (2002). Recent rapid regional climate warming on the Antarctic Peninsula. In *AGU Fall Meeting Abstracts*, volume 1, page 04.

Venter, G. G. (2002). Tails of copulas. In *Proceedings of the Casualty Actuarial Society*, volume 89, pages 68–113.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.

Warburton, J. and Fenn, C. R. (1994). Unusual flood events from an Alpine glacier: observations and deductions on generating mechanisms. *Journal of Glaciology*, 40(134):176–186.

Whitfield, P. H., Bodtker, K., and Cannon, A. J. (2002). Recent variations in seasonality of temperature and precipitation in canada, 1976–95. *International Journal of Climatology*, 22(13):1617–1644.

Willis, I. C., Arnold, N. S., and Brock, B. W. (2002). Effect of snowpack removal on energy balance, melt and runoff in a small supraglacial catchment. *Hydrological Processes*, 16(14):2721–2749.

Wu, J., Wang, X., and Walker, S. G. (2014). Bayesian nonparametric inference for a multivariate copula function. *Methodology and Computing in Applied Probability*, 16(3):747–763.

Xiong, L., Jiang, C., Xu, C.-Y., Yu, K.-x., and Guo, S. (2015). A framework of change-point detection for multivariate hydrological series. *Water Resources Research*, 51(10):8198–8217.

Appendix A

Appendix to Chapter 2

Here is some auxiliary material of Chapter 2.

In this appendix, we explain in detail the proposed MCMC algorithm to sample from the posterior distribution of the model parameters, . Recall that the log likelihood is given by:

(A.1)

(A.2)

(A.3)

(A.4)

We construct a Gibbs sampling scheme where each model parameter is updated separately. Therefore, it is not necessary to compute the whole likelihood for each parameter. In particular, when updating the parameters corresponding to the temperature, , it is only necessary to consider (A.1), (A.3) and (A.4). When updating the discharge parameters, , only (A.2), (A.3) and (A.4) are evaluated. And finally, for updating the copula parameters, , only (A.3) and (A.4) are considered.

The structure of the proposed MCMC method is shown in Algorithm 4. Firstly, it is required to set a vector of initial values for the parameters and the number of MCMC iterations. Then, in each step of the algorithm, each model parameter is updated using a RWMH which is defined in Algorithm 5. Observe that the algorithm is written such that it is not necessary to recalculate the likelihood that was evaluated in previous step for accepted parameters. Finally, Algorithms 1, 2 and 3 separate the computation of the likelihood as the sum of the log-likelihood temperature, discharge and copula, respectively.

These algorithms have been programmed in software R (R Core Team 2013) with the help of the `CDVine` package (Brechmann and Schepsmeier (2013)).

Algorithm 1 Likelihood temperature

Require:
1: **procedure**
2: Calculate marginal distribution parameters
3:
4:
5:
6: **end procedure**

Algorithm 2 Likelihood discharge

Require:
1: **procedure**
2: Calculate marginal distribution parameters
3:
4: Only when
5: **end procedure**

Algorithm 3 Likelihood copula

Require:
1: **procedure**
2: Calculate rank tau parameter
3: Obtain from Depends on the selected copula
4: when and when
5: **end procedure**

Algorithm 4 MCMC algorithm

Require: temperature and discharge series, initial values for , iterations ( ).

```
procedure
    Calculate with the initial values
    Calculate with the initial values
    Calculate likelihood with the initial values
    if l2= then
        Error Message: "Incorrect initial values"
    end if
    Calculate algorithms (1) and (3)
    Calculate algorithms (2) and (3)
    for do
        for do =number of parameters of temperature model
            run RWMH algorithm (5) for temperature parameters
            if new parameter is accepted then
                Update and
            end if
        end for
        for do =number of parameters of discharge model
            run RWMH algorithm (5) for discharge parameters
            if new parameter is accepted then
                Update and
            end if
        end for
        for do =number of parameters of copula model
            run RWMH algorithm (5) for copula parameters

        end for
    end for
    Eliminate burn-in period of chains
end procedure
```

Algorithm 5 Random Walk Metropolis Hastings

```
Require:
procedure
    Simulate
    if                     then                              the number of components in Fourier series
                  mod    (or mod     if               )
    end if
    Construct candidate vector:
    if                        then
                                                                                    algorithm (1)

        Calculate
                                                                                    algorithm (3)

    end if
    if                     then

                                                                                    algorithm (2)
        Calculate
                                                                                    algorithm (3)

    end if
    if                then


                                                                                    algorithm (3)

    end if
    Recover    from previous iteration
    Compute
    if                   then

    else

    end if
end procedure
```

Appendix B

Appendix to Chapter 3

Here is some auxiliary material of Chapter 3.

B.1 Algorithm

In this appendix, we explain the algorithm to obtain the predictive values of the discharge with the conditional probability given in (3.10). Algorithm 6 details the estimation procedure to obtain the predictive mean of the glacier discharge given the temperature, humidity, radiation and precipitation.

For the case that we want to estimate the predictive median of the discharge, we may replace in Algorithm 6 the instructions (6) and (16) by "Compute ——————" and "Compute ——————" respectively.

Finally, for the last prediction method, the conditional probability, is estimated at the beginning of the algorithm and then, it is predicted that if the estimated probability of zero discharge is greater than or obtained with the algorithm if it is smaller.

Algorithm 6 Predictive discharge (using the mean)

Require: Values of meteorological variables: , c-vine copula parameters: and distribution functions:

```
procedure
    if p=0 then
        Compute , and
        Compute and
        Compute
        Simulate
        Obtain the value that verify
        Obtain the value that verify
        Obtain the value that verify
        Obtain
    else
        Compute , , and
        Compute , and
        Compute and
        Compute
        Simulate
        Obtain the value that verify
        Obtain the value that verify
        Obtain the value that verify
        Obtain the value that verify
        Obtain
    end if
end procedure
```

B.2 Density functions as copulas

The joint density functions in (3.1c) and (3.1d) can be expressed in terms of a vine copulas as,

(B.1)

(B.2)

where the superscripts denote the condition of zero or non-zero of discharge and precipitation values: and , and (omitting the superscripts for more clarity)

,

, ,

, , ,

, , , ,

and similar expressions for the other variables

and similar expressions for the other variables

and similar expressions for the other variables

Fig. 3.6 show the structures of the different c-vine copulas used in this paper.

Appendix C

Appendix to Chapter 4

Here is some auxiliary material of Chapter 4.

Algorithm 7 (for c-vine)

Require: data: , c-vine structure: family and copula parameter in each edge

procedure
    for **do**
        Calculate
        Calculate The Kendall correlation
    end for
    for **do**
        Calculate
        Calculate
    end for
    for **do**
        Calculate
        Calculate
    end for So many For-loops as you have
    Set The summary statistic
end procedure

[a]Note the model is composed by 12 c-vine structures of dimension 3, 4 and 5.

Algorithm 8 ABC, with local regression adjustment, for hierarchical model.

Require: Observed data: , c-vine structure[a]: family and copula parameter in each edge, tolerance threshold .

```
procedure
                                              algorithm (7) in every c-vine structure.
                  , where                            M vectors of dimension 12.
                  , where        —                   M vectors of dimension 12.
          where        —                  M matrices of    rows and    columns.
    if family=(Gaussian Copula,Frank Copula) then
              :
    else                                                  Gumbel, Clayton, Joe
              :
    end if                                M matrices of    rows and    columns
          :          where   is in Table 1.3.   Every column of every matrix corresponds
  to a one c-vine structure. M matrices of    rows and    columns.
    for          do

                                              algorithm (7) in every c-vine structure.
                                                          The Euclidean distance.
    end for
    Select                       .
    Compute                                               The Epanechnikov kernel
    Find    and    that minimize                          Local linear regression
    Compute
    Return                      sample of the posterior parameter distribution.
end procedure
```